(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

CARL MURCHISON

GORDON ALLPORT
Harvard University

ARTHUR G. BILLS
University of Cincinnati

E. A. BOTT
University of Toronto

LEONARD CARMICHAEL
Smithsonian Institution

HERBERT S. CONRAD
U. S. Office of Education

ELMER K. CULLER
University of Rochester

JOHN F. DASHIELL
University of North Carolina

ROY M. DORCUS
University of California at Los Angeles

HORACE B. ENGLISH
Ohio State University

FRANKLIN FEARING
University of California at Los Angeles

FRANK A. GELDARD
University of Virginia

CLARENCE H. GRAHAM
Columbia University

ERNEST R. HILGARD
Stanford University

HEINRICH KLÜVER
University of Chicago

CARNEY LANDIS
Columbia University

NORMAN L. MUNN
Bowdoin University

JOHN PAUL NAFE
Washington University

DONALD G. PATERSON
University of Minnesota

LEE EDWARD TRAVIS
University of Southern California

ROBERT C. TRYON
University of California

MICHAEL J. ZIGLER
Wellesley College

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 53

1962

Copyright, 1962, by The Journal Press
Published Quarterly by The Journal Press
Provincetown, Massachusetts, U. S. A.

AUTHOR INDEX

Adams, Henry E. 47
Bendig, A. W. 199
Braun, John R. 453
Buss, Arnold H. 417
Cameron, Jean S. 257
Clausen, Johs. 105
Davie, James S. 161
Dufort, Robert H. 465
Ekman, Gösta 319
Evans, Gloria Carey 457
Evans, Wayne O. 183
Feibleman, James K. 295
Forgays, Donald G. 193
Gogel, Walter C. 91, 475
Gregg, Lee W. 491
Gundlach, Ralph 219
Gustav, Alice 119
Haas, Kurt 409
Hall, George C. 3
Hall, John F. 77
Halpern, Seymour 383
Hawkes, Glenn R. 183, 211
Heimstra, Norman W. 233
Hetrick, W. Robert 409
Ittelson, William H. 245
Jensen, Arthur R. 127
Joy, Robert J. T. 183
Kanfer, Frederick H. 29, 143
Karn, Harry W. 491
Karrer, Rathe 105
King, H. E. 425
Knehr, Charles A. 15
Kodman, Frank Jr. 441
Korn, Sam J. 301
Krinsky, Leonard W. 387
Künnapas, Teodor 319
Loeb, Michael 211
Marston, Albert R. 29, 143
Mayzner, M. S. 155, 469
McBrearty, John F. 143
Molitor, Hazel 193
O'Connell, Donald N. 55
Orne, Martin T. 55
Petersen, Robert L. 289
Phares, E. Jerry 399
Phelan, Joseph G. 281
Pitz, Gordon F. 491
Riess, Bernard F. 219
Rust, Ralph M. 161
Sager, Clifford J. 219
Seidenberg, Bernard 245
Shapiro, Stewart B. 81
Shor, Ronald E. 55
Specht, Priscilla G. 257
Sutton, Rachel S. 37
Teel, Kenneth S. 289
Thumin, Frederick J. 329
Tresselt, M. E. 155, 469
Welch, Livingston 301
Waters, Jerry E. 441
Wells, Harry K. 349
Wendt, G. R. 257
White, Leslie A. 311
Wright, John H. 465

TABLE OF CONTENTS

Conceptual attainment in schizophrenics and nonpsychotics as a function of task structure 3
By George C. Hall

Thinking impairment in acute schizophrenic illness 15
By Charles A. Knehr

The effect of task-relevant information on verbal conditioning 29
By Frederick H. Kanfer and Albert R. Marston

Behavior in the attainment of economic concepts 37
By Rachel S. Sutton

Authoritarianism, statistical rigidity, and a sampling control of response set . 47
By Henry E. Adams

Validation and cross-validation of a scale of self-reported personal experiences which predicts hypnotizability 55
By Ronald E. Shor, Martin T. Orne, and Donald N. O'Connell

Controlled association time as a function of word frequency 77
By John F. Hall

A theory of ego pathology and ego therapy 81
By Stewart B. Shapiro

Convergence as a determiner of perceived absolute size 91
By Walter C. Gogel

Aftereffect of electrical stimulation to the eye 105
By Johs. Clausen and Rathe Karrer

Students' attitudes toward compulsory participation in experiments . . . 119
By Alice Gustav

An empirical theory of the serial-position effect 127
By Arthur R. Jensen

Stimulus discriminability in verbal conditioning 143
By Albert R. Marston, Frederick H. Kanfer, and John F. McBrearty

Incidental learning: a function of associative strength and distance between S-R pairs 155
By M. S. Mayzner and M. E. Tresselt

Studying students as students: a case sample 161
By James S. Davie and Ralph M. Rust

Autonomic effects on estimates of time: evidence for a physiological correlate of temporal experience 183
By Glenn R. Hawkes, Robert J. T. Joy, and Wayne O. Evans

Reinforcement as a function of instructional set 193
By Donald G. Forgays and Hazel Molitor

The Pittsburgh Scales of Social Extraversion-Introversion and Emotionality . 199
By A. W. Bendig

Vigilance for cutaneous and auditory stimuli as a function of intersignal interval and signal strength 211
By Glenn R. Hawkes and Michel Loeb

The uses and interpretations of a follow-up study of patients in pychoanalytic psychotherapy 219
By Bernard F. Riess, Ralph Gundlach, and Clifford J. Sager

Social Influence on the Response to Drugs: I. Amphetamine Sulfate . . . 233
By Norman W. Heimstra

The perception of faces—a further study of the Engel effect 245
By William H. Ittelson and Bernard Seidenberg

Chemical studies of behavior: VI. Placebo and dramamine as methodological controls, and effects on moods, emotions and motivations 257
By G. R. Wendt, Jean S. Cameron, and Priscilla G. Specht

An exploration of some personality correlates to business risk-taking behavior . 281
By Joseph G. Phelan

A study of test-retest results obtained on aircrew proficiency examinations . 289
By Kenneth S. Teel and Robert L. Petersen

The stressed conditioning of psychotics 295
By James K. Feibleman

Conditioned responses: fact and artifact 301
By Sam J. Korn and Livingston Welch

Symboling: a kind of behavior 311
By Leslie A. White

Brightness of monochromatic light in scotopic and photopic vision . . . 319
By Gösta Ekman and Teodor Künnapas

What psychologists should know about marketing and advertising research . 329
By Frederick J. Thumin

Religious attitudes at a small denominational college as compared with Harvard and Radcliffe 349
By Harry K. Wells

Hypno introspection: a contribution to the theory of hypnotherapy. Part I . . 383
By Seymour Halpern

Personal beliefs of schizophrenics and their parents 387
By Leonard W. Krinsky

Perceptual threshold decrements as a function of skill and chance expectancies . 399
By E. Jerry Phares

Some personality correlates of verbal conditioning 409
By W. Robert Hetrick and Kurt Haas

Stimulus generalization with aggressive verbal stimuli: a new paradigm . . 417
By Arnold H. Buss

Anticipatory behavior: temporal matching by normal and psychotic subjects . 425
By H. E. King

Abstract conditioning 441
By Jerry E. Waters and Frank Kodman, Jr.

Stereotypes of the scientist as seen with the Gordon Personal Profile and Gordon Personal Inventory 453
By John R. Braun

The influence of "fake" personality evaluations of self description . . . 457
By Gloria Carey Evans

Food intake as a function of duration of food deprivation 465
By Robert H. Dufort and John H. Wright

Verbal concept attainment: a function of the number and strength of positive instances presented 469
By M. S. Mayzner and M. E. Tresselt

The effect of convergence on perceived size and distance 475
By Walter C. Gogel

The effect of systematic and nonsystematic presentation of stimuli on peripheral visual acuity 491
By Harry W. Karn, Lee W. Gregg, and Gordon F. Pitz

$20.00 per annum
$15.00 per volume
$7.50 per half volume

QUARTERLY
Two volumes per year
Immediate publication

January, 1962
Volume 53, First Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

JANUARY, 1962

(Manuscripts are printed in the order of final acceptance)

Conceptual attainment in schizophrenics and nonpsychotics as a function of task structure 3
By George C. Hall

Thinking impairment in acute schizophrenic illness 15
By Charles A. Knehr

The effect of task-relevant information on verbal conditioning 29
By Frederick H. Kanfer and Albert R. Marston

Behavior in the attainment of economic concepts 37
By Rachel S. Sutton

Authoritarianism, statistical rigidity, and a sampling control of response set . 47
By Henry E. Adams

Validation and cross-validation of a scale of self-reported personal experiences which predicts hypnotizability 55
By Ronald E. Shor, Martin T. Orne, and Donald N. O'Connell

(OVER)

Copyright, 1962, by The Journal Press
Provincetown, Massachusetts, U. S. A.
Entered as second-class matter January 26, 1937, at the post-office at Provincetown, Mass., under the Act of March 3, 1879
Second-class postage paid at Provincetown, Mass.

Controlled association time as a function of word frequency 77
By John F. Hall

A theory of ego pathology and ego therapy 81
By Stewart B. Shapiro

Convergence as a determiner of perceived absolute size 91
By Walter C. Gogel

Aftereffect of electrical stimulation to the key 105
By Johs. Clausen and Rathe Karrer

Students' attitudes toward compulsory participation in experiments . . . 119
By Alice Gustav

An empirical theory of the serial-position effect 127
By Arthur R. Jensen

Stimulus discriminability in verbal conditioning 143
By Albert R. Marston, Frederick H. Kanfer, and John F. McBrearty

Incidental learning: a function of associative strength and distance between S-R pairs 155
By M. S. Mayzner and M. E. Tresselt

Studying students as students: a case sample 161
By James S. Davie and Ralph M. Rust

Autonomic effects on estimates of time: evidence for a physiological correlate of temporal experience 183
By Glenn R. Hawkes, Robert J. T. Joy, and Wayne O. Evans

Reinforcement as a function of instructional set 193
By Donald G. Forgays and Hazel Molitor

The Pittsburgh Scales of Social Extraversion-Introversion and Emotionality . 199
By A. W. Bendig

Vigilance for cutaneous and auditory stimuli as a function of intersignal interval and signal strength 211
By Glenn R. Hawkes and Michel Loeb

The uses and interpretations of a follow-up study of patients in pychoanalytic psychotherapy 219
By Bernard F. Riess, Ralph Gundlach, and Clifford J. Sager

Social Influence on the Response to Drugs: I. Amphetamine Sulfate . . . 233
By Norman W. Heimstra

ERRATUM

In Volume 52, Page 433, Figures 2 and 3 should be as follows:

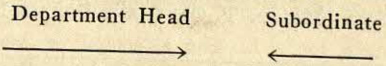


FIGURE 2

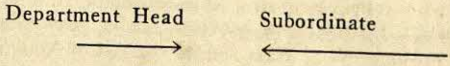


FIGURE 3

This difference between the two figures represents the relative valences of two solutions proposed by the Department Head and by the subordinate, respectively. In Figure 2, the Department Head's idea has higher valence, in Figure 3, the subordinate's has the higher valence.

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 3-13.

CONCEPTUAL ATTAINMENT IN SCHIZOPHRENICS AND NONPSYCHOTICS AS A FUNCTION OF TASK STRUCTURE*[1]

Department of Psychology, Vanderbilt University

GEORGE C. HALL

Observation of schizophrenic behavior in clinic and hospital settings served as the stimulus which led to the investigation reported on these pages. Schizophrenics reputedly suffer from a thinking disorder, yet in interview sessions, testing situations and in their behavior on hospital wards they, at times, seem to show capacity for unimpaired conceptual activity. The following question arose: can the environment be so structured as to maximize the possibility of efficient conceptual function by schizophrenic patients?

Some writers (Apter, 1954; Arieti, 1948) regard schizophrenic thought as a deviation from the normal rather than a unique process. Cameron (1939) has noted that schizophrenics can hypothesize and form concepts but remarks that both hypothesis and concept are frequently inappropriate. Writers (Bleuler, 1951; White, 1948; American Psychiatric Association, 1952) in the clinical area are generally in agreement in regard to the assumption that schizophrenics withdraw from reality in the sense of being less dependent upon external fact for the maintenance of their psychological processes.

From the experimental viewpoint, a review of representative historical studies on nonpsychotics (Davidon, 1952; Heidbreder, Bensley, & Ivy, 1948; Heidbreder & Overstreet, 1948) indicates that various methods of presenting problems requiring conceptual activity can demonstrably influence the conceptual process. Heidbreder (1948, 1949a, 1949b) found that experimental variations in task structure materially affected conceptual order of dominance.

---

* Received in the Editorial Office on August 2, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirement for the degree doctor of philosophy at Vanderbilt University in 1955. Sincere thanks are extended to L. E. Thune, committee chairman, Arthur Canter and Robert Soar who served on the dissertation committee. The writer is also grateful to those at the Murfreesboro Veterans Administration Hospital, Murfreesboro, Tennessee, who cooperated in the execution of the experimental plan. S. E. Abel, M.D., Chief of Professional Services, gave permission for the study to be conducted. Francis H. Deter, Chief, Psychology Service, and his staff, made theoretical and practical contributions.

She concluded that the more "thing-like" the task, the less demand upon the organism to resolve the problem. Greater situational support, apparently, created more favorable conditions for acquisition of concepts. Davidon (1952), too, discovered that his subjects formed concepts more efficiently when the task was structured in terms of greater situational support (pictures) than less situational support (names). It is inferred, then, that normal persons can attain concepts more readily when they are placed in a problem situation structured in terms of the more "thing-like" since this type of task requires less organism activity.

Schizophrenics, too, should react differentially to changes in task structure. This statement is based upon the assumption, evolved above, that the more material available in the environment for solution of the task, the less demand upon the organism solving the problem. In a test situation having less situational support, i.e., fewer available cues, the schizophrenic would necessarily be required to draw upon his own resources in working toward a solution. Using his own method of attaining a concept in the less structured situation, the schizophrenic would, presumably, be more likely to demonstrate the pathological thinking process symptomatic of his disorder. In a conceptual test situation having more cues available there should be less demand upon the schizophrenic's own resources and a greater likelihood of avoiding interference from the pathological thinking process.

The following hypothesis emerges: Schizophrenics' conceptual attainment will approach that of the nonpsychotic persons as the conceptual task becomes increasingly structured.

A. METHOD

The conceptual process requires two subprocesses: First, abstraction; then, generalization. A suitable technique for the study of concept formation is the Heidbreder (1946) type of method. With this method, it is assumed that the subject must abstract, then generalize to attain a concept. This technique is also suitable for structural variation. In the current study there will be a variation in the relationship between the nonsense syllable and its pictorial referent. Not emphasized in previous studies, it is noted that the syllable is transformed from sign to symbol, becoming the common mediating response.

1. *Subjects*

The forty-eight male schizophrenic subjects were drawn from the wards of a Veterans Administration Neuropsychiatric hospital. They were neither acutely ill nor "deteriorated." The test population did not include patients receiving either drugs or psychotherapy. Pilot study results suggested that there would be patients unable to cooperate in the experiment. Inability to cooperate was defined as total rejection of the task at its onset. Patients rejecting the task were not used in the study; data from "mildly disinterested" subjects were included in the statistical analyses. The pilot study had demonstrated that the mildly disinterested patients might well produce a response at any stage of the test session. The patients finally selected for the study were randomly assigned to one of three groups: A, B, and C.

Forty-eight nonpsychotic subjects were selected from the personnel at the hospital. They were closely comparable to the patient population along the dimensions of age and educational level (see Table 1). Each of the nonpsychotic subjects were likewise randomly assigned to one of three groups, A, B, and C, with sixteen in each group.

TABLE 1
COMPARISON OF AVERAGE AGE AND AVERAGE EDUCATIONAL LEVEL FOR SCHIZOPHRENIC AND NONPSYCHOTIC GROUPS UNDER EACH EXPERIMENTAL CONDITION

	Condition					
	A		B		C	
Group	Age	Ed. Level	Age	Ed. Level	Age	Ed. Level
Schizophrenic	35.0	9.6 yrs.	35.0	11.0 yrs.	37.9	10.3 yrs.
Nonpsychotic	35.3	10.5 yrs.	38.5	11.1 yrs.	41.0	11.3 yrs.

2. *Materials*

The materials consisted of three sets of cards, three inches by two and one-half inches: one set of ninety-six pen drawings for all three conditions; one set of six nonsense syllables printed in black on a white background, used in Condition B; one set of six nonsense syllables printed in white on six differently colored backgrounds, used in Condition C.

The pen drawings were divided into three conceptual groups: object, spatial and numerical quantity. There was a further subdivision of the drawings into specific instances: Objects were represented by "face" and "tree" drawings, spatial forms were represented by "circle" and "cross-hatch," numerical quantity by "two" and "six" units. These drawings were arranged into sixteen trials with any single trial containing six pictures, one instance of each concept. Trial one, for example, contained the picture of a face, tree, circle, crosshatch, a picture with two units, a picture with six units. Subsequent trials each contained distinctive, individualized examples of the concept to be attained. The concepts were, essentially, randomly assigned to positions within a series. Measures were taken, however, to avoid having a trial end, for example, with "face" and the next trial begin with "face." Thus, it was hoped that spuriously rapid learning of a concept by virtue of its position with the series would be avoided.

The nonsense syllables and the concept with which each was paired are the following (See Heidbreder in Osgood, 1953, p. 671):

Object	Spatial	Number
RELK—face	FAUD—circle	LING—two
MOFT—tree	PRAN—crosshatch	DILT—six

3. *Experimental Technique*

Condition A. This task was the least "thing-like" of the three conditions. It was the one situation having the fewest cues available.

Instructions: (Schizophrenics) The patient entered the room and efforts were made to establish rapport. After the patient seemed comfortable, the examiner said, "I have something I'd like for you to do for me. I've made up some words and I'd like for you to say them after me so you will get familiar with them." There was a pause. The examiner then said the syllables in random order: "PRAN." The patient pronounced each syllable after the experimenter. If he seemed to have difficulty saying the word, he was helped until he obtained proper pronunciation. The experimenter said the five remaining syllables: "LING," "RELK," "MOFT," "DILT," "FAUD." The presentation order of the syllables was the same for this condition as for Conditions B and C.

Upon completion of the introduction to the syllables by verbalization, the examiner took the first series of pictures and said: "Now, I've given each of these pictures one of the names we've just said together." Trial one, the learning trial, was then begun. The first picture was shown to the subject. The examiner said: "We call this 'LING'." He waited for the patient to say "LING." If the response was not forthcoming, the patient was prompted. The remaining pictures were presented in like manner.

At the end of the first trial, the experimenter said, "Fine, now with these next pictures I want you to give me the name before I tell you what it is." Trial two was presented. Each picture was exposed for five seconds and the patient was prompted if he did not respond at the termination of the time limit.

Throughout the remaining fourteen trials the experimenter responded to the patients' comments as necessary. For example, if the patient protested, "But these are new, I haven't seen them before," the experimenter said, "That's right," and continued with the task. An encouraging attitude was maintained throughout. The patient's critical speculations were not commented upon but correct responses received a substantial verbal reward. This reward appeared necessary to elicit the patient's best efforts.

Instructions: (Nonpsychotic) The experimental technique was the same for nonpsychotics as for schizophrenics.

Condition B. This task was intermediate along the "thing-like" dimension since the nonsense syllables were visually available instead of merely spoken. The syllables were lettered in black on a white three inch by two and one-half inch card.

Instructions: (Schizophrenic) The patient entered the room and efforts were made to establish rapport. After the patient seemed comfortable, the examiner said: "I have something I'd like for you to do for me. I've made up some words . . . ." At this point, the examiner placed the lettered cards before the subject. They read, from the patient's left to right, PRAN, LING, RELK, MOFT, DILT, FAUD. Continuing with the instructions, the examiner said, ". . . I'd like for you to say them after me so you will get familiar with them." There was a pause. The examiner then pointed to the first card and said, "PRAN." The patient pronounced each syllable after the experimenter. If he seemed to have difficulty saying the word, he was helped until he attained proper pronunciation. The remaining syllables were pointed out and pronounced: "LING," "RELK," "MOFT," "DILT," "FAUD."

Following this introduction, the examiner held the first series of pictures and said: "Now, I've given each of these pictures one of the names we've just said together." Trial one was initiated and the first picture presented. The examiner said: "We call this 'LING'." If the response was not forthcoming, the patient was prompted. The remaining pictures were presented in like manner with the examiner pointing to each printed syllable in turn.

Upon completion of trial one the experimenter said: "Fine, now with these next pictures I want you to give me the name before I tell you what it is." Trial two and the remaining fourteen trials were presented without further comment other than necessary prompts and encouragement. Substantial verbal reward was given to correct responses.

Instructions: (Nonpsychotics) The experimental technique was the same for the nonpsychotic group.

Condition C. The final and most "thing-like" of the three conditions, Condition C, added an additional cue to the structure of the task. This cue was color. By printing each syllable on each of the six obviously different colors, it was possible to vary, in accordance with the experimental plan, from subject to subject not only the order in which the syllable appeared but also the color on which it was printed. The instructions and presentation for both schizophrenics and normals of this group are the same as those under Condition B.

4. *Scoring*

Performance was scored in terms of the number of correct anticipations, number of prompts given and the number of concepts attained. The criterion for conceptual attainment was two successive correct anticipations. Trial one was not scored since it was the learning trial. Trial two and the subsequent fourteen trials were scored for each of the six items per trial. Thus, there were ninety basic scores per subject. The number of prompts, correct anticipations and attained concepts varied, of course, from subject to subject.

B. RESULTS

The raw data were analyzed in terms of two variables or dimensions: (1) The comparative effect of task structure upon the number of concepts attained by schizophrenics and nonpsychotics. (2) The effect of varying task structure upon conceptual attainment by schizophrenics as a group and nonpsychotics as a group.

1. *The Effect of Task Structure upon Schizophrenic and Nonpsychotic Conceptual Attainment: A Comparison*

The Mann-Whitney "U" Test (Walker & Lev, 1953) indicates a difference between schizophrenic and nonpsychotic conceptual attainment under all three conditions of task structure. The "U" test analysis deals only with the number of concepts attained, not the type of concept attained.

From Table 2 it is evident that though a difference between the schizophrenic and nonpsychotic group is maintained under Condition B, it is less significant than under Condition A. The nonpsychotic subjects continue to attain more concepts than the schizophrenics.

Condition C yields a z value that is less than that obtained under Condition A, but greater than that under Condition B. Therefore, there is a greater difference in conceptual attainment for nonpsychotic and schizophrenic individuals under Condition A, the least structured of the three conditions. Under each condition, the schizophrenic subjects attained fewer concepts than did the nonpsychotics.

TABLE 2
MANN-WHITNEY "U" TEST APPLIED TO SCHIZOPHRENIC CONCEPTUAL ATTAINMENT AND NONPSYCHOTIC CONCEPTUAL ATTAINMENT FOR EACH OF THE THREE EXPERIMENTAL CONDITIONS: A, B, AND C

Condition	Sum of Ranks for Schizophrenics	Sum of Ranks for Nonpsychotics	Z Value	Sign. Level
A	200.0	328.0	2.41	2.2%
B	221.5	306.5	1.60	11.0%
C	215.0	313.0	1.83	6.3%

2. *The Effect of Task Structure upon Conceptual Attainment within the Schizophrenic Group and within the Nonpsychotic Group*

Values resulting from an analysis of schizophrenic and of nonpsychotic conceptual attainment under the three conditions of task structure are presented in Table 3. The Kurskal-Wallis "H" Test (Walker & Lev, 1953) was applied to: (1) conceptual attainment by schizophrenics under each of the three experimental conditions and (2) conceptual attainment by nonpsychotics under each of the three conditions.

TABLE 3
DIFFERENCES IN CONCEPTUAL ATTAINMENT OF SCHIZOPHRENICS AND OF NONPSYCHOTICS UNDER THE THREE TASK STRUCTURES

	Condition				
	A	B	C		
	Sum of Ranks	Sum of Ranks	Sum of Ranks	H	Sign. Lev.
Schizophrenic	317.5	425.5	433.0	2.65	30%
Nonpsychotic	290.5	373.5	512.0	8.67	2%

Inspection of Table 3 shows an "H" value of 2.65 for the schizophrenic group, significant at the 30 per cent level. The nonpsychotic group data, however, yields an "H" of 8.67 which is significant at the 2 per cent level. It should be noted that the sensitivity of the "H" test is diminished by ties, of which there were many in the raw data. In computing the "H" test, ranking was determined by number of concepts attained. The individual attaining the lowest number of concepts was given the lowest rank. Reference to Table 3 demonstrates that schizophrenics tended to receive lower ranks under Condition A than under either of the other two conditions, which, in turn, were about equal in sum of ranks. Nonpsychotics, on the other hand, showed a progressive increase in the number of concepts attained from Condition A, to B, to C.

C. Discussion

1. *Schizophrenic Conceptual Activity as a Function of Task Structure*

The results of the present study would uphold those who contend that the schizophrenic retains the power to form generalizations. Cameron (1939) maintains that the schizophrenic has the power to hypothesize and generalize. Schizophrenic performance on the Heidbreder type task used in the present study demonstrates the capacity of persons suffering from this disorder to form concepts.

The initial hypothesis, schizophrenics' conceptual attainment will approach that of nonpsychotic persons as the conceptual task becomes increasingly structured, is not supported by quantitative evidence. There are qualitative features which may aid in understanding some of the factors influencing the obtained results: Schizophrenic and nonpsychotic persons showed the greatest difference in conceptual attainment under Condition A, the more abstract of the three conditions. There was less difference between schizophrenic and nonpsychotic performance under Condition B, the intermediate condition along the abstraction dimension. Although performance under Condition C showed a greater disparity between experimental and control groups than under Condition B, the difference was not as great as under Condition A. Inspection of the raw data reveals that five schizophrenics tested in Condition C failed to achieve a concept, one attained three concepts, the remaining ten subjects attained four or more concepts. The absence of the gradation of conceptual attainment, found under both Condition A and Condition B, suggests the possibility of an unknown variable entering into the third condition. It is difficult to account for the abrupt change in conceptual activity under Condition C. Possibly, it could be accounted for by the introduction of color into the structure of the task with a resultant disorganizing affective response. Or, it may simply be that the patients placed in this category who did not respond were unequal to the task on account of their illness.

Examination of the data from another viewpoint, a comparison of schizophrenics' performance under Condition A with schizophrenics' performance under Condition B with schizophrenics' performance under Condition C, demonstrates their reaction to task structure. There would seem to be a "tendency" for schizophrenics to show an increase in conceptual attainment as the structure of the task becomes more concrete. The "tendency" is inferred from the increase in sum of ranks as we move from Condition A to Condition B to Condition C. The differences between conditions, however, are not significant. The nonpsychotics, on the other hand, reacted to task structure with a very definite pattern indicated by their steady increase in number of concepts attained as the task became more structured. Interpretation of this finding is based upon the schizophrenics' withdrawal from the environment; or, the possibility that differences in the actual physical environment do not influence the conceptual behavior of schizophrenics as much as of nonpsychotics.

2. *Nonpsychotic Conceptual Activity as a Function of Task Structure*

Previous "experimental" investigations of conceptual activity have been restricted principally to college student populations. On the other hand, "clinical" studies have emphasized the abnormal. The present investigation includes a nonpsychotic sample more closely approaching general population norms of educational level. Osgood (1953) reports that college level persons consistently performed at the conceptual level on the Vigotsky block problem; those of a lower educational level (noncollege) varied widely, with their mean performance being an intermediate level. The nonpsychotic population used in this study performed conceptually under all three conditions of task structure. An attempt to compare the noncollege group's performance on the Heidbreder type of conceptual task and on the Vigotsky problem is unwarranted because of differences in the structure of the problems. Nor can the nonpsychotic performance in this experiment be directly compared with Heidbreder's college group, since the experimental conditions were not the same. But the ability of the nonpsychotic control group, used in this study, to perform the Heidbreder type of task successfully serves as an important finding for further study. It would be useful to know the difficulty level of the Heidbreder problem compared with other concept formation tasks. It would seem that a series of studies on educational level and its correlation with various conceptual tasks would be suitable empirical evidence for or against the average man's capacity to assume the abstract attitude. Goldstein and Scheerer (1941) maintain that the average person ordinarily does not assume this attitude but can do so with instruction.

This matter of instruction may be broadly interpreted as related to task structure. Reed's (1946a, 1946b) work demonstrated that set could influence conceptual attainment. It appears, furthermore, that not only verbally given set, but task structure given set, in terms of additional cues, can modify conceptual acquisition. The nonpsychotic control group used in the present investigation performed differentially to the three types of task structure ranging from "more abstract" to the "more concrete." They attained fewer concepts under the "more abstract" condition than under the condition which had increased situational support in terms of added cues. The nonpsychotics, then, were capable of utilizing the material perceptually available to them in order to fulfill the demands of the task.

D. Summary

Two groups were selected to test the hypothesis that schizophrenic conceptual attainment would approach nonpsychotic conceptual attainment as the structure of the task became more concrete: one schizophrenic experimental group of forty-eight individuals and one nonpsychotic control group of forty-eight individuals. The latter group was comparable to the psychotics along the dimensions of age and educational level. The members of these groups were randomly assigned to three subgroups representing three differently structured conceptual tasks or experimental conditions: A, B, and C.

The data yielded a difference between schizophrenic and nonpsychotic conceptual attainment under all three conditions of task structure. The most significant difference between the two groups appeared under the least structured experimental condition, A. There was less difference between number of concepts acquired by the schizophrenics and nonpsychotics under the intermediate task structure condition, B. Performance on the final, most structured condition, C, showed a greater difference between number of concepts attained by the two groups than Condition B, but less than Condition A. Under Condition C, the schizophrenics attained either no concepts or three or more concepts. This gap in the distribution suggests that either color, introduced into Condition C's structure, or the severity of the patients' illness may have introduced an uncontrolled element.

Examination of schizophrenic conceptual attainment under Condition A, Condition B, Condition C, reveals an increase in the number of concepts attained as the structure of the task increased. Although there was an increase in the number of concepts with increased structure, it was not statistically significant. The nonpsychotics, on the other hand, also showed a steady increase in number of concepts attained from Condition A to Condition B to Condition C, establishing a statistically significant difference in their performance from condition to condition.

References

1. American Psychiatric Association, Committee on Nomenclature and Statistics. *Diagnostic and statistical manual: mental disorders,* Washington, D.C.: American Psychiatric Association Mental Hospital Service, 1952.

2. Apter, N. S. Research approaches to chronic schizophrenic reactions: methodological problems. *Bull. Menninger Clinic.,* 1954, **18**, 154-161.

3. Arieti, S. Special logic of schizophrenia and other types of autistic thought. *Psychiatry,* 1948, **111**, 325-338.

4. Bleuler, E. The basic symptoms of schizophrenia. In D. Rapaport (*Ed. & Tr.*), *Organization and Pathology of Thought.* New York: Columbia University Press, 1951. Pp. 581-649.

5. Cameron, N. Schizophrenic thinking in a problem solving situation. *J. Ment. Sci.,* 1939, **85**, 1012-1035.

6. Davidon, R. S. The effect of symbols, shifts and manipulation upon the number of concepts attained. *J. Exp. Psychol.,* 1952, **44**, 70-79.

7. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.,* 1941, **53**, No. 2 (Whole No. 239).

8. Heidbreder, Edna. The attainment of concepts: I. Terminology and methodology. *J. Gen. Psychol.,* 1946, **35**, 73-189.

9. ———. The attainment of concepts: VI. Exploratory experiments on conceptualization at perceptual levels. *J. Psychol.,* 1948, **23**, 193-216.

10. ———. The attainment of concepts: VII. Conceptual achievements during card sorting. *J. Psychol.,* 1949, **27**, 3-39. (a)

11. ———. The attainment of concepts: VIII. The conceptualization of verbally indicated instances. *J. Psychol.,* 1949, **27**, 263-309. (b)

12. Heidbreder, Edna, Bensley, M. S., & Ivy, M. The attainment of concepts: IV. Regularities and levels. *J. Psychol.,* 1948, **25**, 299-329.

13. Heidbreder, Edna, & Overstreet, P. The attainment of concepts: V. Critical features and contexts. *J. Psychol.,* 1948, **26**, 45-69.

14. Osgood, C. E. Method and Theory in Experimental Psychology. New York: Oxford University Press, 1953.

15. Reed, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. Exp. Psychol.,* 1946, **36**, 71-87. (a)

16. ———. Learning and retention of concepts: IV. The influence of the complexity of the stimuli. *J. Exp. Psychol.,* 1946, **36**, 252-261. (b)

17. Walker, H., & Lev, J. Statistical Inference. New York: Henry Holt & Company, Inc., 1953.

18. White, R. W. The Abnormal Personality. New York: The Ronald Press, 1948.

VA Hospital
Brockton, Massachusetts

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 15-28.

THINKING IMPAIRMENT IN ACUTE SCHIZOPHRENIC ILLNESS*[1][2]

Department of Psychiatry, Cornell University Medical College

CHARLES A. KNEHR

In the absence of a solution to the problem of the aetiology of schizophrenia or even the question of whether such illnesses present an inherent unity, schizophrenic reactions are differentiated clinically by a set of psychological or behavioral criteria. These behavioral manifestations imply a basic psychological abnormality or impairment reaching seemingly, at times, into every segment of mental life. Examination of the repeatedly described configurations of psychopathological features might suggest some agreement among investigators and clinicians with respect to the postulated nature of the altered underlying psychological functions.

However, a search of the literature reveals that if the role of biochemical, neurophysiological, genetic and other biological factors remains enigmatic, the fundamental nature of the psychological impairment continues to be hardly less of a riddle (Kety, 1959; Medical Research, 1955). Published studies and discussions of psychopathology in schizophrenic patients intimate that the core of the problem resides in an extensive and persisting alteration in the organization of perceptual processes which is manifested, for example, in loss of will, altered capacity to relate to other persons, incongruities of thought and action, looseness of associative connections, loss of the abstract attitude and the appearance of increased perseveration.[3] But, to even casual observation, it is apparent that these and other characteristics do not all appear uniformly or in the same configurations from one patient to another. Laboratory study often fails to verify the impaired abstractive ability or increased perseveration in schizophrenic patients, and frequently they are as readily alerted as the college student. As of now, workers in the field have not arrived at abstracted psychological variables that are invariant from

---

* Received in the Editorial Office on August 4, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] From the Department of Psychiatry, Cornell University Medical College, The New York Hospital (Payne Whitney Psychiatric Clinic), and Hunter College.

[2] The author gratefully acknowledges the skillful technical assistance of Miss Myrtle Guy and Miss Ethel Schwartz in the collection of data, statistical analyses and preparation of results for publication.

[3] Only some of these psychopathological manifestations are ordinarily considered by clinicians to represent impaired thinking.

patient to patient or from one observer to another. Slow progress in understanding these disorders is compounded of many factors, the chief difficulty being this diversity of complex behavioral features. Added difficulties result from the continued use of clinical terminology that implies the older and outmoded faculty approach with its concepts of unitary psychological functions. Investigations have also been without the guidance of an adequate theory of normal perceptual and thinking processes (Hebb, 1949 and 1959; Humphrey, 1951), with the result that studies have been based on a variety of theoretical orientations, and have shown an impressive amount of disagreement. So the question as to how patients with schizophrenic reactions differ psychologically from the normal remains without satisfactory answer.

Even though these psychopathological phenomena have been resistant to experimental probing, it is not unreasonable to hold to the assumption that regularities are to be found. If, instead of working from premises related to the older faculty approach, an investigation is based on what has been learned about cognition in the last fifty years, there is still the promise of moving toward a fuller understanding of schizophrenic thinking disorder.

The first step in the investigation of aberrent thought processes is to describe the extent and functional areas of deviation from normal thinking. This approach is implicit in previous work. Given the proper experimental tools and theoretical orientation, any regularities in the nature of schizophrenic impairment should be discernible. A mature theory of thought would aid in the approach and formulation of an investigation of impaired thinking, but such study may proceed in advance of a complete and adequate theoretical model.

The series of investigations by Spearman and co-workers provided the cue for the present experimental approach. By factorial techniques, Spearman (1927) attempted to describe the domain of cognition by a minimal number of factors or dimensions and asserted at the time that all such factors had been discovered. Although these assertions are open to question, the results of his studies provided a useful set of categories around which a fairly comprehensive battery of tests could be built. If there are four (or even more) general factors in cognitive functioning, tests of the kind most heavily loaded on the stated factors would be expected to bracket a large share of the basic parameters of thinking. Furthermore, an integrative viewpoint would be intended and would involve an assumption often not made explicit in earlier work. Instead of the contingency that atypical thought processes might represent events with laws peculiar to themselves, it is inherent in the dimensional approach that abnormality would denote some deviation along one or more dimensions as a matter of degree.

If it can be assumed that clinical descriptions of psychological deviations in schizophrenics refer to manifestations of persisting alterations in basic thinking processes, then when studied dimensionally, by measurement on the same variables, individuals with impaired thinking should be located differently than normals in the factor space. If there is a common set of modes of thinking deviation, the schizophrenic population should develop a clustering of its own, appearing in some special portion of the factor space, possibly outside the boundaries describing the range of normal cognitive functions.

Experimentally, the issue resolved into a study in which schizophrenic patients would be compared with normals on a battery of tests presumed to be adequately loaded on Spearman's four universal factors plus one non-independent dimension. These are as follows:

1. *G*, or the *general intellective factor*, which is represented functionally as the level of complexity at which abstract eduction of new relations and correlates can be accomplished.

2. The dimension of *span* or constant mental output. This refers to the total capacity for attentive performance and its division when distributed among several mental activities. This factor was found to be independent of retentivity but not of *G*.

3. The factor of inertia or *perseveration* which refers to psychological processes which begin and cease more gradually than their apparent causes.

4. *Oscillation*, or loosely, variability in the sense of changes in the amplitude of performance. As described in the original work, this dimension seems related to some concepts of "mental fatigue."

5. The universal factor of *W* which can be described as *purposive consistency*. While this universal was not taken to be a cognitive factor, its marked influence on cognitive functioning made it important when studying cognition.

To recapitulate, it was hypothesized that if such cognitive dimensions are measurable with a satisfactory degree of reliability and validity, and if schizophrenic thinking disorder represents alterations in these fundamental aspects of mental functioning, then the measurement of these variables by means of a carefully designed test battery should yield a more inclusive picture of schizophrenic deficits in relation to the normal.

A. METHOD

1. *Subjects*

The experimental group was composed of acute schizophrenic patients who were not under treatment with tranquilizers, insulin, or in a course of electroconvulsive therapy (n = 50, mean age 25, $\sigma = 7.27$). All were receiving psychotherapy and the usual daily regimens of socialization, occupational and recreational therapy. All but one of these patients were diagnosed by the same skilled clinician using a consistent set of conservative diagnostic standards.

It was originally intended to match the fifty patients with fifty control subjects. After some interesting data were obtained with the same battery of tests on patients with cerebral organic damage,[4] a dual control group was decided upon; twenty-five to match the younger schizophrenic group, and a similar number to match an older organic group. This would provide the additional advantage of allowing the effects of aging to appear in the control group on the tests to be described.

The control subjects (n = 25) were matched with the schizophrenic group by age (mean age 25, $\sigma = 7.49$), vocabulary achievement, the same proportion of the two sexes, and by a rough approximation to socio-economic status. The loading of the control group with volunteer college students was avoided to a considerable degree even though most of the schizophrenics were intellectually superior, and many were college trained.

2. *Tests*

The tasks were selected to test as nearly as possible the dimensions referred to in the Introduction. In some cases, tests of the kind used in the studies reported by Spearman were duplicated, others were newly designed.

1. Progressive Matrices (1938).[5] This is a British test that has become rather well known in the United States, being a non-language test of the ability to discriminate, to perceive new relationships and to use these relationships in abstract reasoning by analogy. It has been shown to have a high factor loading on *G*. The score was the number of items correctly accomplished without time limit.

2. Wide Range Vocabulary Test (Form B).[5] This is a multiple choice test using one-hundred words of graded difficulty. The score was the number of items correct with no time limit.

3. Flicker Fusion Test Set.[6] This is a self-contained electronic device

---

[4] To be reported elsewhere.

[5] Obtainable from the Psychological Corporation, 304 East 45th Street, New York 17, N.Y.

[6] This instrument is a modification of an electronically controlled flicker apparatus originally designed for Prof. Carney Landis by R. S. Melsheimer of the Bell Laboratories and constructed by Blendinger Instruments of Washington, D.C. The author is most grateful to Prof. Landis for the suggestion for and use of the altered instrument design.

operating a cold cathode light source at frequencies between 20 and 60 cycles per second with continuous monitoring of the frequency by a large calibrated meter. One set of experimental conditions was employed. The threshold was measured at the first appearance of flicker as the frequency declined automatically at a rate of 2 cycles/second on a foveal field subtending 2° at 50 cm. from the eye. Since the light pulse duration was approximately 986 microseconds, the light and dark ratio varied as the frequency declined from 60 to 20 cycles yielding differences in apparent brightness of from 17 to 7 foot-lamberts. These conditions remained constant from run to run. The threshold (CFF) was the mean of twenty trials.

4. Oscillation. The device used is a Pursuit Rotor[7] which was automatically controlled to operate for 15 seconds followed by a rest or off period of 15 seconds, these occurring alternately. A near-instantaneous start and stop tenth of a second synchronous timer recorded the amount of time out of each 15 seconds the subject held the hinged stylus in contact with the target which rotated at a rate of 30 cycles/minute. Testing was for twenty trials with time scores plotted and variability read for trials 3 through 18 from a curve fitted by inspection. Scores were the mean time on target for trials 1-5 and 16-20, and variability the mean of measured differences from the fitted curve.

5. The After-Image test is in reality one point on the dark adaptation curve for red light. The subject was light adapted at an intensity of 154 foot-lamberts for 30 seconds and then was instructed to continue looking in the same direction. At the time the bright adapting light went off, a monochromatic red light (Wratten filter #70, .0167 foot-lambert) began flashing once per second on a white Landolt ring target. The ring subtended slightly less than 2° . Time was recorded for the first recognition of the direction of opening in the ring for three trials following one practice trial.

6. The Series Choice Reaction Time is an automatically controlled sequencing and recording apparatus. The subject stood before a large sloping board on which are mounted three large jewel lights set 20 inches apart in a triangular pattern with three corresponding buttons. In operation, a sequence of twenty reactions was initiated by the experimenter which caused one of the lights to come on. When the subject pressed the button corresponding to the light then on, one of the two remaining lights flashed on while simultaneously the first was extinguished. Each reaction thus set up a new choice situation without delay which precluded the subject's "getting

---

[7] Manufactured by Ralph Gerbrands, Arlington, Massachusetts.

set" for the next trial since he was required to respond immediately to the choice situation which was the result of his just completed response. Each sequence was in different prearranged random order. The reaction times were registered by means of a fifteen channel Esterline-Angus Operation Recorder fitted with a rapid external chart feed. The scores were the mean and variability of the 60 reactions of the last three runs measured from an ink written tape.

7. The Span of Apprehension test is a series of tachistoscopic exposures of random patterns of from 3 to 10 dots. There are three of each stimulus number and are shown as white dots on a dark field. These stimuli were recorded on single frames on motion picture film with appropriately timed ready signals (2 sec.) preceding each 1/30th of a second presentation and followed by a 12 second interval for responding. In Part I, 24 configurations of dots alone were presented in fixed random order. In Part II, 24 similar patterns of dots were presented, but this time with the addition of grid lines in the field. For one third of the stimuli, these lines appeared in the frame before the dots, for another third in the same frame, and for the remainder in the frame following. Responses were plotted against the stimulus number for each subject in both series. The score was the mean amount of under or overestimation of the actual stimulus number presented.

8. Divided Attention. The pursuit rotor was used as one of the two tasks. The other was a head-set and key device so arranged that a 400 cycle tone was fed irregularly by means of punched film programmer to the headset with the subject directed to depress the key with the unused hand only during the time the tone was heard. Three trials of twenty seconds each were given separately with the pursuit task and the auditory following task. The two tasks were then combined for three trials. The amount of loss or gain was calculated for each task separately from measurements of tracings on the recorder chart.

9. Persistence. This apparatus was a device held in the preferred hand with a spring loaded button which the subject was instructed to keep depressed with his thumb for as long as possible. The experimenter acted in such a way as to appear to ignore the subject while he held the button down and the time was being registered on the recorder chart.

10. Opposites. A list of 36 easy to difficult stimulus words were presented orally and the responses and times were recorded manually. The total score was the sum of weighted scores based on difficulty and time.

11. The Minnesota Multiphasic Personality Inventory (MMPI) was added chiefly to screen out controls who might show clinical levels of psychopathol-

ogy and who might have otherwise gone undetected. It was also administered to the patient group.

One might question the failure to include more tests than the ten described above, particularly where a factor analysis might be intended. Two considerations dictated the limitation to the test procedures described. Because patients often exhibit day to day variations in psychopathology, the test battery had to be of such a size that it could be administered in less than a day to gain internal consistency across the several variables being tested. Second, it was considered that excessive testing might interfere with therapy and alter patient's attitude toward treatment and the hospital. The net result was a test battery averaging between three and four hours of actual test time, not counting the MMPI, but requiring six hours when time for lunch and suitable rest periods were added.

B. PROCEDURE

The order of tests remained the same for all subjects. It was assumed that any fatigue effects or diurnal variations would not be significant in amount, especially in view of the variety of test activities and the provision for time out. The Progressive Matrices and Vocabulary were administered to patients in the morning with the remaining eight tests in two forty minute sessions in the afternoon. The MMPI was taken last at the patient's leisure. With availability at a premium in the control subjects, the above schedule was altered where necessary for the members of this group.

C. RESULTS

The essential findings of the study of schizophrenic thinking impairment are condensed into Table 1. At the outset these data can be reduced to the general finding that the acute schizophrenic patients individually and as a group exhibit slight if any impairment on the functions tested when compared to matched controls. It is further asserted that if the abilities measured are unimpaired, then the basic mechanisms involved in these manifestations are probably intact. In other words, the acute schizophrenics of this study appear to have been supplied with the normal complement of neurological equipment which was used for the most part at the same level of efficiency as the controls.

As can be seen from the table, none of the differences between means reaches an acceptable level ($P = .05$) of statistical significance, and the differences are generally small numerically.

Leaving the vocabulary scores aside since these data were used for matching, there was a slight inferiority of abstract conceptual efficiency in the schizophrenics on the Progressive Matrices and the Opposites tests. If the Matrices raw scores are converted, the difference between the groups is approximately six points of I.Q.

TABLE 1

THE MEANS AND STANDARD DEVIATIONS FOR THE CONTROL (N = 25) AND SCHIZOPHRENIC (N = 50) GROUPS WITH THE CRITICAL RATIOS FOR THE DIFFERENCES BETWEEN THE GROUP MEANS

Test	Mean		Standard Deviation		Critical Ratio
	Control	Schizophrenic	Control	Schizophrenic	
Progressive Matrices Raw Score	52.84	49.24	6.64	8.71	1.99
Wide Range Vocabulary Raw Score	81.36	79.12	8.66	11.28	.95
CFF in Cycles/Second	44.87	44.60	4.20	3.65	.28
Variability in CFF Threshold (σ)	.96	.95	.47	.40	1.54
After-Image Duration in Seconds	21.46*	19.45**	14.66	10.57	.60
Series Choice Reaction Time in Seconds	.74	.76	.05	.07	1.86
Variability of Series Choice Reaction Time (σ)	.15	.15	.04	.05	.07
Pursuit Learning—Mean of Trials 1-5 in Seconds	5.68	5.07	1.78	2.36	1.24
Pursuit Learning—Mean of Trials 16-20 in Seconds	8.89	8.86	1.86	2.18	.06
Variability of Pursuit Learning (σ)	.62	.64	.15	.26	.56
Span of Apprehension Response Error—Part I	.50	.78	1.09	.88	1.04
Span of Apprehension Response Error—Part II	—.43	—.08	.95	.84	1.63
Divided Attention Loss in Per Cent	27.70	22.23	18.60	16.43	1.25
Opposites—Raw Score	58.59	53.37**	9.91	13.71	1.87
Persistence—Minutes	4.18	4.42**	2.63	3.53	.33

* N = 24
** N = 49

The results on tests presumed to measure perseveration to some degree in visual and other functions, namely the After-Image Duration, CFF and Series Choice Reaction Time, show differences between the groups too small to require comment.

Various measures of intra-individual variability such as CFF threshold variability, variations in Series Choice reaction times and trial to trial changes, as measured from the fitted learning curve of the Pursuit task, show virtually no differences in favor of one or the other group.

The schizophrenics seemed to do a little, but not significantly, better in resisting loss of performance in the two tasks of Divided Attention. As a matter of interest, some schizophrenic patients showed no loss on this test and were able very successfully to ignore the distracting grid lines of the Span of Apprehension—Part II. It was thought, in viewing these special cases, that perhaps the test might be a measure of dissociative capacity. But the data for the controls showed about the same proportion of normal individuals with such tendencies.

As for persistence, individual differences were spread wide and the group difference favored, but not significantly, better persistence in the schizophrenics.

D. DISCUSSION

The obtained data suggest that this study must be added to those investigations in which "Experimental work has indicated that intellectual deterioration in schizophrenia is more apparent than real, . . ." (Baker and Thorpe, 1956.) The tests which have been considered together in detailing these negative results were grouped on a priori grounds and not on the basis of factor study. As yet, the number of meaningful factors which can be extracted from the battery is not known, although this work is in progress. The study began with the idea that there would be at least five, but as Thurstone (1947, p. 340) stated ". . . a factor analysis does not necessarily reproduce the categories that constitute the preconceptions of the investigator." Since these schizophrenic patients performed essentially in the same manner as the controls, it is reasonable to infer that the schizophrenics will not differ along the dimensions to be extracted and that a factor analysis will serve mainly to describe the test battery without contributing essentially to the stated findings.

Previous work in this laboratory led to the expectation that no significant differences would be found in tests of perseveration (Brown, 1956; Knehr and Brown, 1957) and possibly those of abstract conceptual functions. From the results of other studies and our assessment experiences, it seemed that any changes to be observed in the schizophrenic groups would be more likely to occur in variability, persistence and attentional functions. In relation to attention in the sense of alerting or readiness, the finding that acute schizophrenic patients performed like controls in an earlier study by Knehr (1954) did not seem to alter the possibility of finding differences on the tests of the present investigation. That the schizophrenics of this study showed the same capabilities and performances as the controls in all functions tested was unforeseen and is at variance with much published work. The problem is to account for such negative results.

Perhaps the inconsistency between these and other data on impairment in schizophrenics is due in part to the nature of the patients studied. An older schizophrenic patient population usually indicates many years of chronicity in a narrowly restricted environment, the only one with which the patient may interact. These patients are the result of unknown selective processes insofar as it is they who become chronic and often show little spontaneity in surroundings which provide lack of adequate therapeutic help or incentives to return along the road to adjustment, where that might be possible. The high incidence of cerebral pathology in one such group reported recently (Howie, 1960) suggests further that psychopathology and reported test results may frequently represent more the consequences of cerebral organic factors than the schizophrenic process.

A study by Hall (1951) based upon the work of Goldstein, Gelb and Vigotsky, assumed as others have done that impairment in schizophrenics could be observed if only the problem were addressed by means of the right measures. On seven tests of abstraction, some acute schizophrenics showed little or no impairment, either quantitative or qualitative, while chronic cases were shown to be similar in many ways to organics. Hall holds that, in using quantitative procedures, one may be overlooking possible subjective factors which may show up only in loose, unstructured perceptual situations. It seems quite possible that he merely obtained impaired scores where brain damage might have been a factor, while normal scores were obtained in individuals with intact nervous systems. Lehrmann (1940) suggested earlier that if deterioration is observed in schizophrenia, organic etiology must be involved. There was no evidence for cerebral anomalies or damage in the patients of the present study and they had not been subjected to excessive environmental restriction except as might result largely from their own behavior. They were relatively young and at the time they were tested would not be considered chronically ill. The obtained results tend to support these assertions.

The question of test sensitivity arises when clinical observation leads to the expectation of some measurable variations only to find the obtained differences either nonexistent or too small to be clearly related to the particular illness-control dichotomy. This may be a fictitious problem. A patient may perform more poorly than the limits of his ability, but could hardly be expected to perform better than his present limitations will allow. If he performs at levels not significantly below those of matched controls, whatever the tests, we must conclude that on these tested functions he does not have added limitations, his ceiling is not appreciably lowered relative to the controls, and the tested functions are at least as efficient as his actual performances.

So, while it would be difficult to prove, it seems reasonable to assume that the test measures selected for the study reported here were capable of measuring impairment if altered basic functioning were a part of the clinical picture. Stated in other terms, if the acute schizophrenic is as concretized in his conceptual activity, as paleologic in his thoughts and as desymbolized and desocialized as Arieti (1959) indicates, then these tests ought to be adequately sensitive.

If, as suggested, our tests were adequate to the measurement of altered psychological functions, then there is no sufficient evidence of any general or specific type of impairment in our schizophrenic patient group on the functions tested. To borrow terminology from the computing engineer, the data processing systems of these patients appear intact. The problems were fed into systems which, for all we knew in advance, might be programmed differently from a "normal" system. Our results showed that the "right" answers showed up on the read-out using "normal" systems as standards of comparison. This suggests that not only was the circuitry intact, but that the programming appeared similar to that for the normal system.

Since the schizophrenics of this study appear to be unimpaired in the sense of some general deficiency in the neural mechanism or neural organization, it is possible that the clinically observed psychopathology may represent altered organization under special circumstances. When compared to the normal, perceptual activities are no less complex and may vary in inclusiveness by degree from patient to patient. This notion is not different from the idea that two equally bright individuals, as measured by a test heavily saturated on *G*, may be found to think and respond quite differently on an emotionally charged political issue.

For some years since the early studies of honesty by Hartshorne, May and Shuttleworth (1930) much effort has been made to provide a theoretical antidote to the ideas of specificity in behavior engendered by those investigations. Interest in generalization has led to a minimizing of the role of specificity in behavior. But when the term motivation is used in the more usual sense of the "why" of particular behavior sequences, the concept of specificity is at the very core. Motivationally, one does not get behavior in general although response equivalence to a particular behavior outcome is often characteristic. This is one of the reasons why a statement about a patient's level of introversion does not allow a predictive statement to be made about the next behavior. In fact, the further from behavior specificities our abstractions lead us, the less predictive they become.

In a series of studies at Fels Institute, Lacey and Lacey (1958) have been able to demonstrate highly individual physiological patterns to stressful situations, the same pattern to a variety of stressors. If functions closely allied to autonomic activity can be organized into such highly individualized configurations, it may be expected that motivational tendencies can be organized by the normal processes of learning in innumerable and relatively specific ways, requiring only the appropriate stimulus configuration to set classes of complex central activities in motion. In the schizophrenic, a large share of these may be, from the standpoint of their behavioral outcomes, similar to those of other persons, while some, resulting from unusual conditions of reinforcement, will appear inappropriate, strange, unusual or to occur in strange (to the observer) sequences. When statistically frequent or impressive enough in other respects, they might constitute the syndrome of the illness.

This discussion has been leading to the assertion that there may be no sound psychological reason for assuming that there is a basic inconsistency in not finding specific or generalized cognitive impairment in patients exhibiting motivational and emotional anomalies, or that these anomalies should have effects that regularly diffuse into unrelated activities. The intact brain is capable of innumerable kinds of organization from person to person and within the same head. And in the presence of such varied organizations, there is no reason to suppose a necessarily concomitant variation in inertia, the limits of sensory data handling capacity per unit time, or the general level of complexity at which conceptual activity can take place. When the schizophrenic is identified as such, it may be on the basis of relatively specific and unusual developments within the area of motivation and emotional responsiveness. Although impairment was observed in a study by Malamud and Palmer (1938), the losses were ascribed to motivational factors, especially a ". . . disturbance of activities requiring the directional control of thought." The present discussion implies that such disturbances may be observed under some circumstances and not in others.

Maier's rats, when subjected to "conflict" situations from which there was no escape, would develop anxiety and stereotypy in behavior that was extremely resistant to retraining (Maier, 1949). We are told that the odd behaviors produced in Liddell's sheep with similar procedures were not generally observed out of the situation which had been employed to produce the deviant behavior. In these animals, there was no evidence for a general impairment or total alteration in mental organization. The abnormality was rather an alteration in some of many functions.

If it can be further verified that acute schizophrenic illnesses do not represent generalized cognitive impairment, the possibility of discovering the sources and manner of producing the aberrant behavior and its reversal remains bright. Under these circumstances, it would be difficult to see how biochemical and genetic factors could play more than a subsidiary role whereas psychological studies of learning, motivation, conflict, and emotion may hold the eventual key to the problem with a consequent increase in therapeutic success.

E. SUMMARY

A group of acute schizophrenic patients was compared with a matched control group on a battery of ten tests based upon Spearman's early work on the factors of cognition.

The schizophrenic patients showed little or no impairment relative to the controls, none of the observed differences reaching a satisfactory level of statistical significance.

The data provide further confirmation for some opinions and results of previous studies which suggest that acute schizophrenic patients are not generally impaired in such presumed capacities as the perception of new relationships, eduction of correlates, concentration and division of attention, discrimination and the learning of a sensori-motor skill. They were equally persistent and no more variable or perseverative than control individuals.

It is concluded that psychopathology in these patients does not preclude normal functioning in other respects, and that it need not imply significantly impaired perceptual or thinking processes in general.

REFERENCES

1. ARIETI, S. (Ed.) American Handbook of Psychiatry. Vol. I. New York: Basic Books, 1959.
2. BAKER, A. A., & THORPE, J. G. Some simple measures of schizophrenic deterioration. *J. Ment. Sci.,* 1956, **102**, 838-846.
3. BROWN, N. L. A study of sensory, motor and ideational perseveration in schizophrenic and diffuse brain damaged patients. University of Pennsylvania Ph.D. Thesis, 1956. Mic. 56-2352, Microfilm Library of Congress, Washington, D.C.
4. HALL, K. R. L. The testing of abstraction, with special reference to impairment in schizophrenia. *Brit. J. Med. Psychol.,* 1951, **24**, 118-131.
5. HARTSHORNE, H., MAY, M. A., & SHUTTLEWORTH, F. K. Studies in the Organization of Character. New York: Macmillan, 1930.
6. HEBB, D. O. The Organization of Behavior. New York: Wiley, 1949.

7. ———. Intelligence, brain function and the theory of mind. *Brain,* 1959, **82**, 260-275.
8. Howie, D. L. Some pathological findings in schizophrenia. *Amer. J. Psychiat.,* 1960, **117**, 59-62.
9. Humphrey, G. Thinking. New York: Wiley, 1951.
10. Kety, S. S. Biochemical theories of schizophrenia. *Science,* 1959, **129**, 1528-1532, and 1590-1596.
11. Knehr, C. A. Schizophrenic reaction time responses to variable preparatory intervals. *Amer. J. Psychiat.,* 1954, **110**, 585-588.
12. Knehr, C. A., & Brown, N. L. Perseveration in schizophrenic and brain damaged patients. *J. Psychol.,* 1957, **43**, 269-275.
13. Lacey, J. I., & Lacey, B. C. Verification and extension of the principle of autonomic response stereotypy. *Amer. J. Psychol.,* 1958, **71**, 50-73.
14. Lehrmann, S. R. Schizophrenic "deterioration." *Psychiat. Quart.,* 1940, **14**, 140-156.
15. Maier, N. R. F. Frustration. New York: McGraw-Hill, 1949.
16. Malamud, W., & Palmer, E. M. Intellectual deterioration in the psychoses. *Arch. Neuro. Psychiat.,* 1938, **39**, 68-81.
17. Medical Research: A Midcentury Survey. Vol. II. Boston: Little, Brown, 1955.
18. Spearman, C. The Abilities of Man. New York: Macmillan, 1927.
19. Thurstone, L. L. Multiple Factor Analysis. Chicago: Univ. of Chicago Press, 1947.

Department of Psychiatry
Cornell University Medical College
1300 York Avenue
New York 21, New York

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 29-36.

THE EFFECT OF TASK-RELEVANT INFORMATION ON VERBAL CONDITIONING*[1]

Department of Psychology, Purdue University

FREDERICK H. KANFER AND ALBERT R. MARSTON

A. PROBLEM

Reports of studies in verbal conditioning have indicated that the acquisition of verbal responses under these procedures is a function of a number of variables which have previously not been considered as crucial determinants of human verbal learning. For example, response acquisition was shown to be a function of the characteristics of the experimenter (Ferguson & Buss, 1960), of the personality variables in subjects drawn from college populations (Sapolsky, 1960), and of the immediately preceding experience of the subject (Kanfer and Karas, 1959). In a recent study (Kanfer and Marston, 1961) it was demonstrated that variation in the ambiguity of several aspects of the verbal conditioning task may affect rate of learning and transfer. An analysis of the verbal conditioning paradigm suggested that the relative lack of specific information about the task may represent one factor which enhances variability among subjects by allowing usually less important variables to become dominant. Furthermore, it was suggested that the verbal conditioning procedure represents several simultaneous tasks to *S*s. Among these is the simple problem of discovering the purpose of the experiment and the relationship between *E*'s reinforcing stimuli and *S*'s own responses.

The present experiment examined further the effects of varying task-relevant information on acquisition rate, and on transfer to a new list in a verbal conditioning procedure. It was hypothesized that increase in information concerning *S*'s particular task may change the task from a relatively complex problem-solving task to a simple discrimination. Therefore, it was argued, increase in task-relevant information should result in more effective learning while irrelevant information should retard acquisition of the response.

---

* Received in the Editorial Office on August 7, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This research was supported in part by Research Grant M-3700 to the first author, from the National Institute of Mental Health, United States Public Health Service.

B. METHOD

1. *Subjects*

*S*s were 36 male and 24 female college students recruited from undergraduate psychology classes. They were randomly assigned to four experimental groups and each group contained the same proportion of male and female *S*s.

2. *Apparatus and Procedure*

The learning tasks consisted of a modified verbal conditioning task similar to one first used by Taffel (1955). Each *S* was shown a pair of words consisting of one mildly hostile verb (rated in an earlier study by Buss and Durkee, 1958), paired with a neutral word. The stimuli were typed on 4 × 5 inch cards and were presented through an opening in a masonite board which separated *S* and *E* in a darkened experimental room. Each word pair constituted one trial. All words were equated for frequency of usage, using Thorndike and Lorge's (1944) word count.

All *S*s were asked to select one of the two presented words as each pair appeared in the window. Twenty stimulus pairs were presented in different random order, correct responses varying right and left, for each of six blocks of twenty trials (120 trials). During the first block (pre-acquisition) *E* made no response. On trials 21-120 *E* said "good" in a flat tone after *S* used a hostile word. Following the 120th trial *S* was told that he would be doing the same thing with another list of words. A second list was then presented using stimulus pairs which were equated with the first list for hostile connotation. Reinforcement was again given on trials 21 to 120. The sequence of the two lists of word stimuli was randomly determined for each *S*.

After presentation of the second list an awareness questionnaire was given to all *S*s. Answers to seven questions concerning *S*'s ability to describe the purpose of the experiment and the response reinforcement contingency were transcribed verbatim and rated on a 6 point scale for degree of awareness. The criteria and procedure for rating are described in an earlier study (Kanfer and Marston, 1961). Reliability of ratings was established by independent ratings made by both authors. A product-moment r of .96 was obtained.

3. *Experimental Design*

The amount of task-relevant information differentiated four groups by the nature of the instructions which they received for the task described above. Thus the groups differed only with regard to the content of the instructions. Task irrelevant or misleading information was given to one group (Misinformation-group) by referring to the pronunciation of the stimulus words, and in the same context stating "good means that you are correct." A second group (No Information-group) received the usual verbal conditioning instructions which mentioned neither the reinforcing stimulus nor the nature of the task. A third group (Points-group) was given some information concerning the general purpose of the task and the nature of the reward. They were told "good means that you have scored a point; score as many points as you can." Finally, the fourth group (Information-group) was given additional information concerning the meaning of the reinforcing stimulus and its relationship to the task by the statement "good means that you have selected the correct word."

C. Results and Discussion

An analysis of variance on the first pre-acquisition block for the four groups indicated that no differences existed among the groups in number of correct responses emitted prior to administration of the reinforcing stimulus $(F < 1.0)$. This analysis indicates that the instructions did not affect the initial selection of the two classes of stimulus words.

The mean numbers of correct responses during acquisition are presented in Figure 1. It can be seen that the amount of learning increases with the increase of information from group to group. A trend analysis on these data is summarized in Table 1. The statistical analyses indicate that the trends suggested by Figure 1 are highly significant with regard to the over-all amount of learning and the rate at which acquisition took place. The over-all means of correct reponses over all blocks were: Misinformation: 60.33; No-Information: 65.20; Points: 79.20; and Information: 91.53. Using Lindquist's method of critical differences (Lindquist, 1953, p. 93) it was found that all groups significantly differ from each other at $p < .05$, except the Misinformation and the No-Information groups. An analysis of the linear trend components similarly shows significant differences among groups,

TABLE 1
SUMMARY OF TREND ANALYSIS ON LIST 1[1]

	Over-all Means		Linear Trend		Quadratic Trend		Cubic Trend	
	MS	F	MS	F	MS	F	MS	F
Treatments	2993.29	11.43*	1.41	8.87*	.282	1.87	.029	1.07
*S*s in groups	261.96		.159		.151		.027	

[1] Degrees of freedom for all F tests = 3 and 56
* $p < .001$

with the fastest rate of acquisition in the group receiving most task-relevant information (Information-group), and the Misinformation-group showing the slowest rate of acquisition. T-tests of linear trends against zero show that all groups except the Misinformation-group have slopes which significantly differ from zero. Thus, all groups except the Misinformation group manifest some improvement in performance. As indicated in Table 1, the quadratic component does not significantly differ for the groups. How-

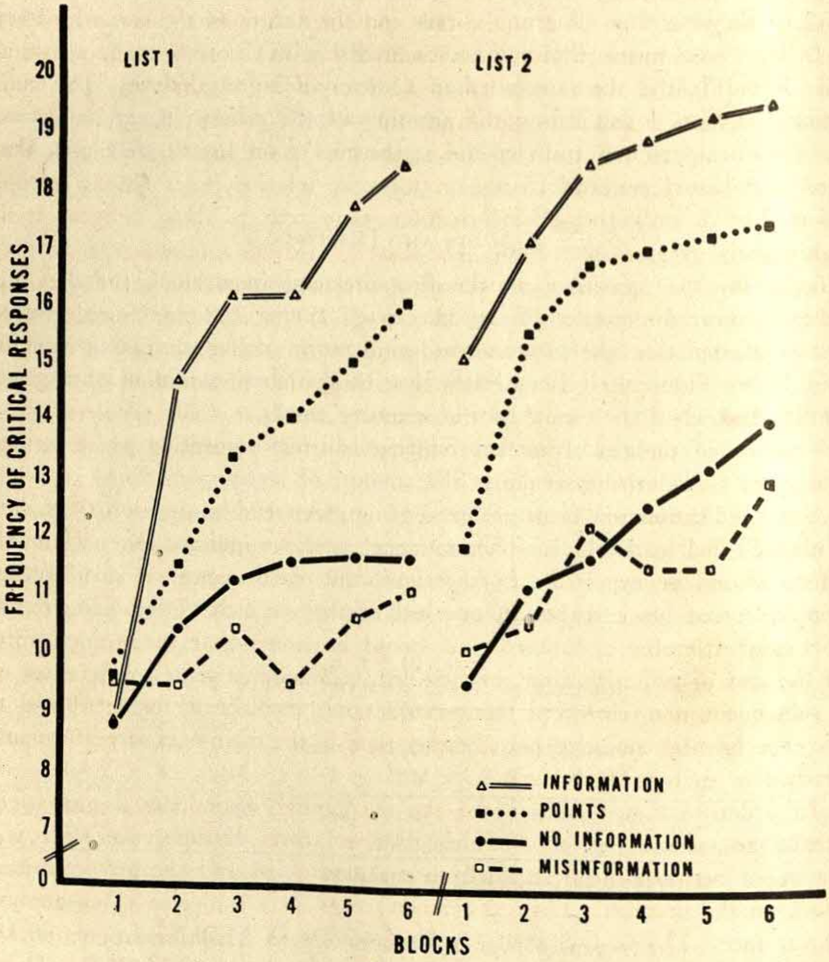


FIGURE 1
Mean Response Frequencies Per Block Over Both Lists

ever, when the total quadratic component is tested against zero, a t value of 23.0 is obtained, indicating that all groups show a similar quadratic trend. Similarly, an F of 1.07 for the cubic component shows that the groups do not differ significantly with regard to this component. However, the cubic component for all groups, when tested against zero, yields a significant t of 2.10 showing that all group curves contained a similar cubic component.

In order to compare learning on a second task with that on initial acquisition, List 1 and List 2 were included in an over-all trend analysis. The comparison of Lists 1 and 2 is significant for over-all means ($F = 26.09$) but not for linear or quadratic trends, i.e., the curves for List 2 are higher than those for List 1 without change in form or relative order among groups. As in List 1, only the Misinformation-group fails to show a linear trend significantly greater than zero. The List × Group interaction is not significant for the measures of over-all means and quadratic trends. A significant linear interaction ($F = 3.66$, $p < .05$) was found. Simple effects tests indicated that the group slopes were more similar on List 2 than on List 1 (see Figure 1). These findings indicate that presentation of a second similar task eliminates some of the necessity for *S* to solve problems about the nature of the task; thereby resulting in greater learning in all groups and greater similarity in slopes.

Some indication of "what is learned" is given by the degree of transfer. Unless *S*s had learned to discriminate the correct response class, no transfer effects would be expected. Furthermore, the performance on a subsequent non-reinforced block represents one test of the specificity of learning to the particular stimulus pairs used. It should be noted that the groups differ at the end of acquisition on the first list. Consequently, any differences on a subsequent non-reinforced block could not unequivocally be attributed to the experimental condition per se rather than to the final level of performance reached by each of the groups.

In order to take advantage of the availability of individual pre-experimental response biases in obtaining transfer scores, a difference score was computed between the first block on the second list and the pre-acquisition block on the first list. These difference scores were subjected to an analysis of variance. The analysis yields an F of 4.78 with 3 and 56 d.f., significant at $p < .01$. The mean difference scores are: Information: 6.33; Points: 2.20; No-Information: 0.67; and Misinformation: 0.60. The group which was given maximal task-relevant information in this study thus shows a performance increment of considerably greater magnitude than any of the other groups. The No-Information and Misinformation groups, in fact, reveal so little change from their initial performance on List 1 that one strongly suspects a qualitative difference in the learning of these groups when contrasted with the Information-group. Since neither the No-Information nor the Misinformation group transfer scores are significantly different from zero (as tested by t tests), these data lend general support to the hypotheses that a reduction in task-relevant information results in *S*'s learning of specific responses rather than the general characteristics of a discriminable response class.

The awareness ratings of all groups were subjected to a simple analysis of variance. The resultant F of 3.29 ($p < .05$), indicates differential effects of the experimental instructions on *S*'s ability to verbalize the response-reinforcement contingency. To some extent, the additional task-relevant information in some of the groups would in itself be expected to yield an increased ability to specify the response-reinforcement contingency. The mean awareness ratings were: Information: 4.87; Points: 3.80; No-Information: 3.73; and Misinformation: 2.73. Lindquist's (1953) method of obtaining a locus of the critical differences shows that only the Information group significantly differed from the Misinformation-group. For all other comparisons among groups the critical difference values do not reach significance at $p = .05$.

It should be noted that the Points and No-Information groups correspond to groups run in an earlier study (Kanfer and Marston, 1961). Comparisons of these corresponding groups with t-tests on acquisition trends, transfer scores, and awareness ratings produced no significant differences. The findings indicate apparent stability in the results using these procedures. With a somewhat different task in which *S*s gave three-place digits, Tatz (1960) varied task information and similarly found that informed groups were superior to uninformed groups.

The results support the hypothesis that increasing task-relevant information in the verbal conditioning procedure used here leads to increased learning. The data further substantiate the suggestion that the usual verbal conditioning procedures present *S* with several concurrent and interdependent tasks, thereby reducing learning of the critical response class and introducing a variety of other variables which dictate each *S*'s individual approach to this complex task.

D. SUMMARY

The present experiment investigated the effects of varying task-relevant information, controlled through instructions to the *S*, upon learning in a

verbal conditioning procedure. Four groups of *S*s received the following levels of task information: (a) the misleading implication that Good referred to correct pronunciation (Misinformation group); (b) the usual verbal conditioning instructions with no information about the task or the reinforcing stimulus (No-Information group); (c) the additional information that the reinforcing stimulus (Good) meant that a point had been scored (Points group); and (d) the information that Good meant that a correct response had been made (Information group).

All *S*s were given word pairs consisting of a neutral word and a mildly hostile word. On each trial *S* was asked to select one of each pair. After 20 non-reinforced trials, E said "Good" whenever *S* selected a hostile word. One hundred acquisition trials were given on the initial list of pairs (20 non-reinforced). On a second list, matched for degree of hostile connotation, 20 non-reinforced trials were followed by 100 acquisition trials.

Analyses of learning trends, amount of transfer on the initial non-reinforced block of the second list, and awareness ratings indicated that the Information group showed significantly greater learning, a steeper learning curve, greater transfer, and higher awareness ratings than the groups receiving less or misleading information. On the second list acquisition was similar to performance on the first list, except for an increase in the height of the learning curves of all groups and an increased similarity in group slopes. The data supported the contention that task-relevant information facilitates learning in verbal conditioning experiments.

REFERENCES

1. Buss, A. H., & Durkee, A. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *J. Consult. Psychol.,* 1958, **22**, 415-418.
2. Ferguson, D. C., & Buss, A. H. Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. *J. Consult. Psychol.,* 1960, **24**, 324-327.
3. Kanfer, F. H., & Karas, S. Prior experimenter-subject interaction and verbal conditioning. *Psychol. Rep.,* 1959, **5**, 345-353.
4. Kanfer, F. H., & Marston, A. R. Verbal conditioning, ambiguity, and psychotherapy. *Psychol. Rep.,* 1961 (In Press).
5. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton, 1953.
6. Sapolsky, A. Effect of interpersonal relationship upon verbal conditioning. *J. Abnorm. & Soc. Psychol.,* 1960, **60**, 241-246.
7. Taffel, C. Anxiety and the conditioning of verbal behavior. *J. Abnorm. & Soc. Psychol.,* 1955, **51**, 496-501.
8. Tatz, S. J. Symbolic activity in "Learning without awareness." *Amer. J. Psychol.,* 1960, **73**, 239-247.

9. THORNDIKE, E. L., & LORGE, I. The Teacher's Book of 30,000 Words. New York: Bureau of Publications, Teachers' College, Columbia Univ., 1944.

Department of Psychology
Purdue University
Lafayette, Indiana

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 37-46.

BEHAVIOR IN THE ATTAINMENT OF ECONOMIC CONCEPTS*

College of Education, University of Georgia

Rachel S. Sutton

A. Introduction

Failure to distinguish between process and product in concept formation has led to confusion. It is essential that data be obtained on the growth and development of economic concepts and that they be analyzed to determine the process of development peculiar to them. There is evidence to support the belief that the development of social concepts (Danziger, 1958; Miller and Horn, 1955) follows paths of its own that do not fit in with the theoretical model elaborated in connection with physical concepts (Navarra, 1955; Piaget, 1960; Russell, 1956).

Economics, an area that deals with the satisfaction of human wants, offers an unusual opportunity for creative imagination and cue searching for attributes that place an experience or object in one category or another. It is a relatively undeveloped area in the elementary school. It has no history of being taught as a subject nor of being presented logically (Gavian, 1942; Bienvenu, 1958). The elementary school child learns only a causal way from home and school about the powerful forces that propel the economy forward or pull it backward.

Capital accumulation in the American economy takes place by a two-sided process of saving money, or withholding current money income from expenditure on current consumption, and investing current money income in a stock of valuable goods to produce more capital. A person may save because of the egotistical desire for prestige and power that the possession of capital confers, in order to improve the living standards and position of the family, or to share the gains of civilization with the human race.

B. Theory

Theory and research on the attainment of economic concepts has attempted to answer such questions as these:

1. How do children achieve the information necessary for isolating and learning a concept?

---

* Received in the Editorial Office on August 8, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

2. How do children retain the information gained from encounters with possibly relevant events so that it may be useful later?
3. How is retained information transformed so that it may be rendered useful for testing an hypothesis still unborn at the moment of first encountering new information?
4. What are the general features, if any, in the growth of economic concepts?
5. To what extent are the responses of children a cultural product of the environment?

1. *How do children achieve the information necessary for isolating and learning a concept?*

In the literature describing first-language acquisitions (McCarthy, 1954) there is much to indicate that children easily form large abstract categories. There are numerous cases in which the child over-generalizes the use of a conventional word. The name given a thing by an adult for a child is determined by the frequency with which various names have been applied to such things in the experience of the particular adult. Each name is at the utility level (Brown, 1958) for the group that uses it. The child's vocabulary is more immediately determined by the naming practices of adults. Each thing is given its most common name, which categorizes in a maximally useful way. With some hierarchies of vocabulary the more concrete terms are learned before the abstract. The most abstract terms are never learned first. A hierarchy develops in both directions from a middle level of abstraction. The sequence in which words are acquired is set by adults rather than by children and may ultimately be determined by the utility of the various categorizations.

2. *How do children retain the information gained from encounters with possibly relevant events so that it may be useful later?*

Research in concept development does not show how children manage the discriminations, the inclusions and the exclusions involved in learning the denotative range of certain terms. Nor does it show the kind of abstractions and generalizations that enter into the thinking of a person of any age as he grasps the significance of various kinds of terms in a situation of social communication.

The effective strategies for attaining concepts depend upon the use of an initial focus. One learns concepts by the association of external stimuli with internal mediating stimuli by a simple law of frequency or contiguity. Conceptions or standard meanings are instruments of identification, supplementation, placing an object in a system, and grouping facts so that general laws or conclusions may be drawn from them. At every stage of development each experience should lead up to a certain amount of conceptualizing of impressions and ideas. Without conceptualizing or intellectualizing nothing is gained that can be carried over to a better understanding of new experience. Each experience should leave behind an increment of meaning, a better understanding of a subsequent situation, a clearer future plan and purpose of action. Meanings are working tools for further apprehensions. The analysis that results in giving an idea the solidity and definiteness of a concept is emphasis upon that which gives a clue for dealing with the unknown.

All cognitive activity (Bruner, *et al.,* 1956) involves and is dependent on the process of categorizing. The working definition of a concept is the network of inferences that are or may be set into play by an act of categorization. The concept is a network of significant inferences by which one goes beyond a set of observed critical properties exhibited by an object or event to the class identity of the object or event in question, and thence to additional inferences about unobserved properties of the object or event. Categorizing serves to cut down the diversity of objects and events that must be dealt with uniquely by an organism of limited capacities, and it makes possible the sorting of functionally significant groupings.

The greater the discriminatory capacity of the organism, the more reduced and implicit can become the detachable reactions finally included in the stable mediation process. The place of the organism in the evolutionary scale (Osgood, 1953) is indicated by the quality of discriminations it can usually make. The more mature and intelligent the human individual the less overt his symbolic processes, as exemplified by the hosts of fine discriminations that characterize his language behavior.

In an analysis of a child's understanding of money (Schuessler and Strauss, 1950) the concept, "money," was not grasped immediately in its entirety; the development of comprehension was gradual, the sign behavior of the child became almost imperceptibly more elaborate, systematic, complex, flexible, and non-egocentric.

3. *How is the retained information transformed so that it may be rendered useful for testing an hypothesis still unborn at the moment of first encountering new information?*

To the degree that a distant end controls a sequence of inquiries and observations and binds them together as means to an end curiosity assumes a definitely intellectual character. The growth, structure, and dynamic quality of a child's conceptual framework is not a mysteriously revealed body of absolute truth but rather a process by which truth is approached.

An essential condition for concept formation is the association between a common response and a variety of stimuli. Conceptions attain a miximum of definite individuality and generality in the degree to which they show how things depend upon one another or influence one another. Control of observation, imagination, and memory so as to select and give proper weight to data as evidence depends upon the possession of a store of standardized meanings or conceptions. They are not ends in themselves but instrumentalities for further understanding and aids to the interpretation of the obscure. The ultimate educative value of all cognitive processes is measured by the degree to which they become working tools in the creation and development of new experience.

Imagination is the ability to make the absent present, and it is this ability which leads men to the stage of the hypothesis in their acts of thought. It arises in relationship to experience and goes beyond experience. The quality and range of experience individuals have had is a determining factor in releasing or restricting their imaginative flights and in restricting or releasing their ability to think (Hullfish and Smith, 1961). It is the relentless effort to see beyond the immediate and the actual, coupled with responsible checking out of ideas which emerge that the creative work of the world is done. Imaginative incidents and episodes that have a certain coherence hang together on a continuous thread and lie between flights of fancy and measures employed to establish a conclusion. When an evaluative dimension appears in the process, an attitude is being formed (Rhine, 1958).

The key to effective thinking is the controlled use of sentiency, memory, imagination, and categorization in a balance that is appropriate to a particular purpose or problem at hand. Thinking is good or bad only in relation to its directing purpose. The key to wisdom lies in the selection of worthy purposes. Generalization results from the process of enlarging and enriching experience so that adequacy of meaning may be tested (Bode, 1940) in contexts of expanding comprehensiveness. As patterns of meaning are enlarged and reconstructed (Dewey, 1933) generalizations emerge to which beliefs may be grounded and judgments controlled. Good sense is good judgment, a power of discernment about what is to be done and what is to be abstained from in the ordinary affairs of life.

4. *What are the general features, if any, in the growth of economic concepts?*

In a study of children's earliest conceptions of economic relationships

Danziger (1958) reported that most of the children appeared to be at a lower stage in their understanding of economic relationships in production than exchange. This was not surprising when one considers that the children have first-hand experience in acts of purchase, but depend on second-hand evidence for their ideas about production and the work situation. The stages of conceptual development do not owe their appearance simply to internal maturation, but depend to a large measure on the intellectual materials with which the child is provided.

Danziger (1958) observed that four stages occur in the development of economic concepts:

(a) An initial pre-categorical stage occurs when the child lacks economic categories of thought altogether. There is no special realm of economic concepts differentiated from social concepts in general.
(b) At the second, or categorical, stage the child's concepts appear to represent a reality in terms of isolated acts which are explained by a moral or voluntaristic imperative.
(c) At the third stage the child becomes able to conceptualize relationships as such by virtue of the fact that a reciprocity is established between previously isolated acts. These relationships are in turn isolated and cannot be explained in terms of other relationships.
(d) The isolated relationships become linked to each other to form a system of relations. There is a conceptualization of a totality wherein each part derives its significance from its position in the whole.

5. *To what extent are the concepts of children a cultural product of the environment?*

Category learning is one of the principal means by which a growing member of a society is socialized, for the categories that one is taught and comes to use habitually reflect the demand of the culture in which they arise. Within a given society the core values of the culture perform an analytic function with respect to structuring value experiences. A creative aspect of concept attainment behavior is that the patterning of decisions does reflect the demand of the situation in which the person finds himself. The primary stage is one of cause and effect and the latter stages are influenced by cultural determinants (Navarra, 1955).

Cue searching is an attempt to find attributes one can use as a basis for deciding whether a particular object should be placed in one category or another. Confidence can be increased by a tendency to see a present task as similar to a past task on which one has used a cue successfully. Cultural differences and individual differences in venturesomeness, with respect to the use of cues, doubtless exist as well as how much information a person requires and how much risk he is willing to take.

C. PROCEDURE

Whether there is sufficient external stimuli for elementary school children to develop categories of economic thought is debatable. In an effort to answer the question eighty-five children were chosen from 425 in six typical public schools located in northeast Georgia. The children were arranged alphabetically in grades one through six and every fifth child was chosen to respond to questions about the use of capital and means of supplying it. It was assumed that the systematic questioning of a child would supply information about the way concepts are formed in the area under investigation. The questions and conditions under which they were asked were the same for each child.

1. Do you like to earn money?
2. How do people get money?
3. What will money buy?
4. What does money do for a person?
5. What can be used in place of money?
6. What is a bank?
7. Why do people like to have money in a bank?
8. What happens to money in a bank?
9. Are there other places to put money?
10. Why do they draw it out?
11. How do people put money in a bank?
12. Why do people save?

The children were thoroughly familiar with the tape recorder and well-acquainted with the investigator from previous visits to the schools so that their replies were natural, free, and spontaneous. At no time was there a preliminary discussion of the questions used in the interview. Replies to the questions were later transcribed for analysis.

Information was collected about each child regarding his father's occupation, mother's occupation, number of siblings in the home, and the child's ability to learn as measured by his mental age and achievement age. Replies to each question were analyzed by grades and categories of economic thought.

D. RESULTS

Language was a significant factor in the replies to each question by younger and older children. For example, to the question, "What will money buy?" high average second grade children replied, "Food, clothes, lots of things";

low average, "Candy, drinks, bubble gum." High average fourth grade children replied, "It will buy houses, useful things like clothes, shoes, and things you need"; low average, "Clothes and food." High average sixth grade children replied, "It will buy everything one needs, everything but happiness"; low average sixth grade, "Clothes, shoes, caps—just about everything."

Responses to the question, "What happens to money in a bank?" also indicated the child's facility with language or lack of it. High average second grade children answered, "You keep it there so you won't spend it." "It is safe and won't get stolen away." "It comes to a larger amount of money"; low average second grade, "So you will have money." High average fourth grade said, "If you keep it there a certain amount of time, you get an added amount of money"; low average, "They put it in a bank and put it in a safe." High average sixth grade said, "It collects interest"; "It is locked in a vault and kept safe"; low average, "It just stays there until you need it for something."

To the question, "Do you like to earn money?" the first, second, and third grades answered in the affirmative. The older children made qualifying statements, such as, "If it is something I like to do"; or "If it is for something I want to buy." One child in a sixth grade mentioned getting an allowance and not earning; others referred to earning by performing chores at home.

To the question, "How do people get money?" most of the replies were in terms of work and different types of work, such as, "By selling things," "Picking cotton," "Working for others," "Out of the bank," "From Mommy and Daddy," "By stealing and cheating," "By work in a business for two weeks or one week at a time."

"What will money buy?" resulted in a long list of responses. Younger children thought of immediate desires such as toys, dolls, candy, a trip, drinks, clothes, and food. Older children mentioned "Food," "Clothes," "Houses," "Cars," "Hospitals," "Apartments," "Airplanes," "Swimming pools," "Everything," "Everything but happiness."

To the question, "What does money do for a person?" younger children replied, "It helps one to stay alive, to pay taxes, and to go places." Older children continued to moralize, "Makes one greedy and selfish if used wrongly." "It helps some people and hurts others." "It helps to pay for sickness, accidents, and old age."

First grade children replied to the question, "What can be used in the place of money?" "Checks"; "You just have to do without." A second grade child said, "In place of money we can use our heavenly Father. I wouldn't like to be rich, but I'd like to have a little money to spend." A fourth grade child said, "Charge it." Sixth grade children mentioned, "Gold," "Silver," "Tobacco, cotton," "Trading cars and other valuables," "Happiness and kindness"; "Use friendship and love, and be very kind to everybody, and somebody might help you sometime."

To the first grade children a bank was a "Piggy bank," "A place to put money." To older children a bank was "A place where you keep money." "It's a place that if you have money that you want to keep in case of an emergency you can put it in a bank and keep it unless you want to have a check cashed."

People like to have money in a bank, "So they won't lose it"; "Burglars can't steal it"; "To pay the taxes": "They won't have too much in a pocketbook"; "It gives them security"; "It helps prepare for the future, for example, if you are eleven or twelve years old and you have money in the bank, you are saving for a college education."

When money is put in a bank "Santa Claus gets it when he comes"; "It keeps good"; "It is put in a vault; you do not see it again unless you draw it out"; "It stays there until you need it and want to use it for something"; "If you keep it there long enough, you earn interest, three per cent or four per cent"; "They count how much you have and give you a bank book, and they put it in a safe in a little sack with your name on it, I guess."

The younger children mentioned several places to put money other than a bank: "A billfold," "A piggy bank," "Your pocket," "A sugar bowl," "A drawer," "Behind a picture on the wall," "A cookie jar." Older children used the same ideas in longer expressions. "Lots of people don't believe in banks, and they hide it." "You spend it, and it's gone." "You can put it in a chest in your home or you can put it in a wall vault that you have at your house." One child in a sixth grade suggested, "They can buy stock."

People draw money out, "To buy something," "For a birthday," "To pay bills"; "They want to make a down payment on a new house"; "To buy a car if they don't have enough money at home"; "To pay for food, furniture, or hospital bills." People put money in a bank, "By mailing it," "By putting it through a slot," "By opening an account"; "They tell a cashier that they want to put in some money; he puts it in a vault and makes a record of it."

People save, "To have for Christmas," "For a rainy day," "To buy things," "To help poor people," "To buy trucks, cars, and go places," "For a college education"; "When people get old, they won't be able to work"; "The more you save the more you can buy"; "They save because they think it is right to save"; "Some of them save to get a lot of money, and some save just to be saving."

One thousand twenty replies of eighty-five children were arranged in six categories:

1. No response, for example, "I don't know." One per cent of the replies were in this category.
2. A pre-categorical stage where a thing is named with apparently little understanding of economic meaning, for example, "A bank is a place to keep money." In this category there were 63 per cent of all replies.
3. The category of goodness or badness, rightness or wrongness, moral or immoral, irrespective of economic function, for example, "Money will buy everything but happiness." Eighteen per cent of all replies were in this category.
4. Two isolated acts or factors which are in no way connected with economic significance, for example, "Some of them save, just to be saving." Twelve per cent fell in this category.
5. Two acts involving a reciprocity which cannot be explained in terms of other economic relationships, for example, "If you keep it there a certain amount of time, you get an added amount of money." In this category there were 5 per cent of all replies.
6. The subjective explanation gives way to the objective; the single act derives its significance from its position in a system of relationships and is no longer conceived of in an isolated way, for example, "A person may save money and invest it to earn more." One per cent of the replies were in this category.

The children in this investigation were at an initial stage of development in their understanding of money and the accumulation of capital. To them money was a thing to spend for the near-at-hand and the immediate. A bank was a safe place to keep money until it was needed for spending. Older children were slightly less ego-centric than younger ones. The majority of replies were on the pre-categorical level. Age, intelligence, and socio-economic background had little effect on the understanding of the production credit of money. Older, more intelligent children tended to moralize and explain in terms of rightness and wrongness. There was little evidence that the respondents had been stimulated by their environment to explore the creative uses of capital and the numerous ways by which it could be acquired. One child in the sample, an eight-year-old, with average ability, whose mother is an employee of a bank, answered the question, "Why do people save?" with the response, "To earn more."

E. SUMMARY

A sample of eighty-five children between the ages of six and thirteen in public schools of northeast Georgia were asked a series of twelve standard questions regarding money and the accumulation of capital. The replies were recorded on tape, transcribed, and later analyzed for ways of forming categories of economic thought. The result that a majority of replies, irrespective of age, intelligence, and socio-economic backgrounds, were in the beginning stages of conceptualization suggested a need for more external stimuli in the development of economic concepts, and raised a question about generality of the condition.

REFERENCES

1. BIENVENU, H. J. Economic education: problems and programs. *Elem. Sch. J.*, 1958, **59**, 97-104.
2. BODE, B. H. How We Learn. Boston: Heath, 1940.
3. BROWN, R. How shall a thing be called? *Psychol. Rev.*, 1958, **65**, 14-21.
4. BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A Study of Thinking. New York: Wiley, 1956.
5. DANZIGER, K. Children's earliest conceptions of economic relationships. *J. Soc. Psychol.*, 1958, **47**, 231-240.
6. DEWEY, J. How We Think. Boston: Heath, 1933.
7. GAVIAN, R. W. Education for economic competence in grades 1 to 6. New York: Bureau of Publications, T. C., Columbia, 1942.
8. HULLFISH, H. G., & SMITH, P. G. Reflective Thinking: the Method of Education. New York: Dodd, 1961.
9. MCCARTHY, D. Language development in children. *In* L. Carmichael (*Ed.*), *Manual of Child Psychology*. 2nd. ed. New York: Wiley, 1954.
10. MILLER, L. N., & HORN, T. Children's concepts regarding debt. *Elem. Sch. J.*, 1955, **55**, 406-12.
11. NAVARRA, J. G. Development of scientific concepts in a young child. New York: Bureau of Publications, T. C., Columbia, 1955.
12. OSGOOD, C. E. Method and Theory in Experimental Psychology. New York: Oxford, 1953.
13. PIAGET, J. Child's Conception of Geometry. London: Routledge and Kegan Paul, 1960.
14. RHINE, R. J. Concept-formation approach to attitude acquisition. *Psychol. Rev.*, 1958, **65**, 363-370.
15. RUSSELL, D. H. Children's Thinking. Boston: Ginn, 1956.
16. SCHUESSLER, K. F., & STRAUS, A. L. Study of concept learning by scale and analysis. *Amer. Sociol. Rev.*, 1950, **15**, 752-762.

College of Education
University of Georgia
Athens, Georgia

Published as a separate and in *The Journal of Psychology,* 1961, **53**, 47-53.

AUTHORITARIANISM, STATISTICAL RIGIDITY, AND A SAMPLING CONTROL OF RESPONSE SET*

Department of Psychology, Louisiana State University

HENRY E. ADAMS[1]

A. INTRODUCTION AND PROBLEM

Recent investigations of the California F scale (Adorno, Frenkel-Brunswik, Levinson, and Sanford, 1950) have indicated that the F scale appears to be contaminated with response set (Bass, 1955, Cohn, 1952). F-scale items are constructed so that agreement is scored as an indication of authoritarian attitudes. Reversed F scales composed of the original F-scale items phrased negatively rather than positively have been constructed by various investigators (Bass, 1955; Chapman and Campbell, 1957; Jackson and Messick, 1957; Messick and Frederiksen, 1958). These latter studies with the reversed F scale have reported a small negative correlation between F and reversed F scores which has been attributed to acquiescence set in both scales.

On the basis of such evidence, three different positions with respect to the significance of acquiescence have been taken. Bass (1955) has indicated that acquiescence appears to be a well-defined personality trait not identical to authoritarianism. Couch and Keniston (1960) have supported this position and presented data indicating some of the basic personality characteristics of "yeasayers" and "naysayers." However, as the authors note, some of the personality scales used as criteria were contaminated by the variable, *i.e.,* acquiescence, which they were investigating. Gage, Leavitt, and Stone (1957) have adopted the view that acquiescence is actually a manifestation of authoritarianism. This approach is consistent with the viewpoint of Leavitt, Hax, and Roche (1955) who have indicated that the F scale confounds two related variables, form and content. However, Leavitt and his colleagues have noted that this contamination is in the right direction since authoritarians tend to agree and are submissive; thus the discriminatory power of the F scales is actually increased. After considering the same evidence, Messick and Frederiksen (1958) have suggested that the demon-

---

* Received in the Editorial Office on August 14, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to express his gratitude to I. A. Berg and R. N. Vidulich for their encouragement and criticism of this study.

strated acquiescence set should be corrected, i.e., either utilized as a suppressor variable or statistically removed, thereby increasing the validity of the F scale. Statistical procedure for accomplishing these corrections have been suggested by Webster (1958) and Helmstadler (1957).

One of the crucial variables governing the appearance of response set, as Berg (1959) and Cronbach (1946) have indicated, is the ambiguity or lack of structure of the item. F-scale items appear to possess ambiguity for some subjects (*S*s) but not for others. Thus certain individuals seem to respond on the basis of set while others respond on the basis of item content. The individual who responds to both form and content is probably rare. Cronbach recognized this tendency when he observed that the ability of the subject to give structure to test items is a crucial variable to be controlled in eliminating response sets.

Two heterogeneous groups involved in F-scale studies may thus account for the disparate results, one responding to item content and the other on the basis of response set. Response set should be regarded as an error score for one group of *S*s and as a true score for the other group. Thus applying statistical correction procedures to the total group would create problems of reliability and validity. In terms of the present study, a sampling control in which the group is divided into *S*s who respond to item content in one case or with a response set in the other would be more appropriate. Cronbach (1946) indicated a similar approach when he suggested discarding the data obtained from individuals with strong response biases. Then the scores on the F and reverse F scales would reveal four groups: an authoritarian group, an equalitarian group, *S*s with a negative response set (negativism), and *S*s with a positive response set (acquiescence). The authoritarian and equalitarian groups would thus be responding to item content, while the acquiescent and negativistic groups would be responding with a response set.

Gage, Leavitt, and Stone (1957) and Leavitt, Hax, and Roche (1955) have suggested that acquiescence is consistent with authoritarian attitudes and have presented evidence in support of this hypothesis. An appropriate test of their position would determine whether or not the personality characteristics of the authoritarian are similar to those of acquiescent individuals who respond to a personality scale which is relatively independent of acquiescence. In the present study, rigidity has been selected for investigation since is has been suggested (Adorno, Frenkel-Brunswik, Levinson, and Sanford, 1950) that rigidity is one of the foremost characteristics of authoritarianism and a technique not contaminated by acquiescence is available for measuring it. When confounding of content and response set is eliminated by sampling control, thus, it is equally possible that rigidity is associated with negativism rather than with acquiescence. For example, Asch (1958) and Barnes (1956a, 1956b) using Berg's (1957) Deviation Hypothesis have presented evidence that negative response sets are associated with neuroses. In view of these findings, the hypotheses of the present study are: (a) The authoritarian person is more rigid that the equalitarian when both are responding to item content; (b) Individuals with negative response sets are more rigid than individuals with acquiescent response sets; (c) Individuals with authoritarian attitudes are more rigid than individuals who exhibit acquiescent response sets; (d) When a sampling control for the confounding of response set and item content is provided, reliable content and response set scales may be developed.

B. METHOD

Subjects were 48 students from an introductory psychology class at Louisiana State University who were volunteers and available. The age range was from 18 to 36 years with a mean age of 22.3 years. They were given the California F scale, a reverse F scale, and the Perceptual Reaction Test (PRT) (Berg, Hunt, and Barnes, 1949), the latter providing a measure of statistical rigidity (Adams, 1960). In order to obtain a measure of rigidity, the Like Much, Like Slightly, Dislike Slightly, and Dislike Much options of the PRT were arbitrarily assigned values from one to four, respectively. The variance for each *S* was computed using single item responses separately as scores. Thus, the variance of the test indicates how much *S* varied in his selection of these options from item to item. The variance was then multiplied by 100 to eliminate decimals and is called a variance score (VS). The theoretical range of these VSs is from 0 to 225. A zero score would require marking the same option for every item. The reliability of VSs ranges from .88 to .97 with different samples (Adams, 1960). The F scale used consisted of 15 of the original F scale items stated in the positive direction while the reverse F scale consisted of the remaining original 15 items stated in the negative direction. A description of these scales and the items is given by Messick and Frederiksen (1958). This version of the F and reverse F scale was selected because the reliability of both scales was found to be very low when corrected statistically for response set (Messick and Frederiksen, 1958). Thus it is possible with these scales to determine whether a sampling control of response set will increase reliability for the content and for response set scales since true and error scores are not confounded.

On the basis of F and reverse F scales, the 48 *S*s were divided into four groups, i.e., authoritarian *S*s who were high scorers on the F scale and low on the reverse F scale; acquiescent *S*s who were high on the F scale and the reverse F scale; equalitarian *S*s who were low on the F scale and high on the reverse F scale; and negativistic *S*s who were low on the F and reverse F scales. Whether an *S* achieved a high or low score on a scale was determined by comparing his score with the median of the scale. Variance scores (PRT rigidity scores) (VS) were computed for each of these groups.

C. Results and Discussion

As shown in Table 1, a single classification analysis of variance yields an F test at the .05 level of significance when groups are compared on the basis of VS. The differences between means, as indicated in Table 2, were computed using one-tail t-tests since the hypotheses were directional. The negativistic group was found to be significantly more rigid than the acquiescent group ($p = .05$), the authoritarian group significantly more rigid than the equalitarian group ($p = .05$), and the negativistic group significantly more rigid than the equalitarian group ($p = .01$).

Correlations were computed between the F and reverse F scales using the total group of *S*s. The F and reverse F scales were also correlated for those *S*s who consistently responded to content (authoritarian and equalitarian *S*s) and for those *S*s who responded primarily on the basis of set (the negative and acquiescent *S*s). As revealed in Table 3, the Pearson product moment correlations between F and reverse F were —.22 ($p =$ ns.) for the total group, —.77 ($p = .01$) for the item content group, and .57 ($p = .01$) for the response set group.

TABLE 1
ANALYSIS OF VARIANCE OF THE DIFFERENCES BETWEEN GROUPS WITH RESPECT TO VARIANCE SCORES

Source	df	Mean Square	F	p
Groups	3	2034.3	2.99	.05
Within	44	680.2		

These results indicate that the authoritarian is more rigid than the equalitarian when the influence of response set is controlled by sampling techniques. Rigidity as measured statistically in the present study does appear to be associated with authoritarian attitudes. However, the negativistic group is significantly more rigid than the acquiescent group. This supports Asch's (1958) and Barne's (1956a, 1956b) findings which indicate that neuroses are more frequently observed among persons who tend to respond negatively. While the difference in degree of rigidity between the acquiescent group and the authoritarian group in the present study is not statistically significant, it is in the direction hypothesized. Thus, while not conclusive, doubt is cast on Gage, Leavitt, and Stone's (1957) and Leavitt, Hax, and Roche's (1955) hypothesis that acquiescence is highly related to authoritarianism.

TABLE 2

MEANS AND DIFFERENCES BETWEEN MEANS ON VARIANCE SCORES

Group		Means		Differences between Means			
				a	A	E	N
Acquiescence	(a)	103.9			9.8	11.3	19.1*
Authoritarian	(A)	94.1	a			21.1*	9.3
Equalitarian	(E)	115.2	A				30.4**
Negativism	(N)	84.8	E				

* Significant at .05 level as determined by t-tests.

** Significant at .01 level as determined by t-tests.

TABLE 3

CORRELATIONS BETWEEN F AND REVERSE F SCALES[1, 2]

Group	N	Correlation	Spearman-Brown Corrections	p
Total Group	48	—.22		n.s.
Content group only	24	—.77	—.87	.01
Response Set group only	24	.57	.73	.01

[1] Scored in the Response Set direction

[2] Pearson product moment correlations

The correlation between the F and reverse F scales for the total group is quite close to the one obtained by Bass (1955). However, when response set is controlled by a sampling technique, a large increase in the size of the correlations is apparent, despite the fact that only 30 items are involved. When Spearman-Brown corrections are applied, as shown in Table 3, these correlations can be considered to be at an acceptable level of reliability. Accordingly, it is probably more profitable to correct for response set on an individual rather than a statistical basis for the total group. In the case of the F scale, response set is an individual phenomenon which some *S*s exhibit while others do not. The appropriate control to use if one is primarily interested in investigating authoritarian attitudes would seem to be the elimination of those *S*s who respond chiefly on the basis of response sets. This form of control attempts to hold constant the *S*'s ability to attribute structure to stimuli in general and to item content in particular.

D. SUMMARY

In the original investigation of the F scale, it was demonstrated that one important characteristic of the authoritarian person was rigidity. However, the *S*s used in these studies had been responding either to item content or response set or both. The hypotheses of the present study are: (a) The authoritarian is more rigid than the equalitarian and acquiescent *S*s; (b) *S*s with negative response sets are more rigid than acquiescent *S*s; (c) when the confounding of response set and item content is controlled by a sampling technique, reliable response set and content scales can be developed.

Forty-eight *S*s from an introductory psychology class were given the California F scale, a reverse F scale, and the Perceptual Reaction Test which yields a measure of statistical rigidity. On the basis of the F and reverse scores, *S*s were divided into four groups—authoritarian, equalitarian, acquiescent, and negativistic. With the exception of the authoritarian *S*s who were not significantly more rigid than the acquiescent *S*s, the hypotheses were all supported by the data. Suggestions were made for the control of response set by attempting to hold constant the *S*s' ability to attribute structure to F scale items.

REFERENCES

1. ADAMS, H. E. Statistical rigidity in schizophrenic and normal groups measured with auditory and visual stimuli. *Psychol. Rep.*, 1960, **7**, 119-122.
2. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950.
3. ASCH, M. J. Negative response bias and personality adjustment. *J. Counsel. Psychol.*, 1958, **5**, 206-210.
4. BARNES, E. H. Response bias and the MMPI. *J. Consult. Psychol.*, 1956a, **20**, 371-374.
5. ———. Factors, response bias, and the MMPI. *J. Consult. Psychol.*, 1956b, **20**, 419-421.
6. BASS, B. M. Authoritarianism or acquiescence? *J. Abnorm. & Soc. Psychol.*, 1955, **51**, 616-623.
7. BERG, I. A. Deviant responses and deviant people; the formulation of the Deviation Hypothesis. *J. Counsel. Psychol.*, 1957, **4**, 154-161.
8. ———. The unimportance of test item content. *In* B. M. Bass & I. A. Berg (*Eds.*) *Objective Approaches to Personality Assessment*. New York: Van Nostrand, 1959. Pp. 83-99.
9. BERG, I. A., HUNT, W. A., & BARNES, E. H. The Perceptual Reaction Test. Evanston, Ill.: 1949.
10. CHAPMAN, L. J., & CAMPBELL, D. T. Response set in the F scale. *J. Abnorm. & Soc. Psychol.*, 1957, **54**, 129-132.
11. COHN, T. S. Is the F scale indirect? *J. Abnorm. & Soc. Psychol.*, 1952, **47**, 732.
12. COUCH, A., & KENISTON, K. Yeasayers and naysayers: agreeing response set as a personality variable. *J. Abnorm. & Soc. Psychol.*, 1960, **60**, 151-174.

13. CRONBACH, L. J. Response sets and test validity. *Educ. Psychol. Measmt.*, 1946, **10**, 3-31.
14. GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescence set for authoritarianism. *J. Abnorm. & Soc. Psychol.*, 1957, **55**, 98-103.
15. HELMSTADLER, G. C. Procedure for obtaining separate set and content components of a test score. *Psychometrika,* 1957, **22**, 381-393.
16. JACKSON, D. N., & MESSICK, S. A note on ethnocentrism and acquiescent response set. *J. Abnorm. & Soc. Psychol.*, 1957, **54**, 132-134.
17. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215-221.
18. MESSICK, S., & FREDERIKSEN, N. Ability, acquiescence, and "authoritarianism." *Psychol. Rep.*, 1958, **4**, 687-697.
19. WEBSTER, H. Correcting personality scales for response sets or suppression effects. *Psychol. Bull.*, 1958, **55**, 62-64.

Department of Psychology
Louisiana State University
Baton Rouge 3, Louisiana

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 55-75.

VALIDATION AND CROSS-VALIDATION OF A SCALE OF SELF-REPORTED PERSONAL EXPERIENCES WHICH PREDICTS HYPNOTIZABILITY*[1]

Studies in Hypnosis Project, Harvard Medical School and Massachusetts Mental Health Center[2]

RONALD E. SHOR,[3] MARTIN T. ORNE, AND DONALD N. O'CONNELL

A. INTRODUCTION

A number of theoretical formulations on the nature of hypnosis have focused upon the subject's relative lack of concern with external reality and his proneness for experiences where external reality is not the major determinant of subjective reality. Many concepts have been used in reference to this phenomenon; for example, dissociation (Janet, *et al.,* circa 1900), altered state of the person (White, 1941), breakdown and fusions of ego boundaries (Kubie and Margolin, 1944), alterations in ego functioning (Gill and Brenman, 1959).

Recent formulations by Shor (1959; 1962a) attempt to account for this phenomenon under the concept of *trance.* Trance is defined as the extent to which the usual waking orientation to generalized reality has faded into the more distant background of awareness so that ongoing experiences are relatively isolated from the usual kind of critical self-appraisal. Trance is viewed as one of three major constituents of hypnosis and as a major constituent of many related states.

An important deduction from these formulations is the prediction that most individuals who can readily become profound hypnotic subjects have had many profound "hypnotic-like" experiences which have occurred naturally in the normal course of living. The theory supposes that these individuals have the ability to suspend their usual generalized reality-orientation so that "hypnotic-like" experiences can occur. This hypothesized ability has been

---

* Received in the Editorial Office on August 29, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was supported in part by contract AF49 (638)-728 from the Air Force Office of Scientific Research, and a grant from the Human Ecology Fund.

[2] We wish to thank our co-workers—Emily F. Carota, Esther Helfman, Jacqueline Mehler-Amati, and Karl Scheibe—for their critical comments and editorial assistance in the preparation of this manuscript. Much appreciation is also due in this regard to Mrs. Elizabeth Scherer.

[3] Postdoctoral Fellow, MF-7860-C3, National Institute of Mental Health, Public Health Service.

termed *tranceability* which is a component of, but is distinct from, *hypnotizability*. In other words, it is hypothesized that at least one permanent attribute of mental functioning lies behind the ability to achieve profound hypnosis. Such an attribute, conceived as a cognitive ability, is viewed as cutting across the currently common classifications of personality traits, such as hysteria or submissiveness.[4]

In no sense, however, do the writers believe that this attribute is the *only* determinant of hypnotizability. Undoubtedly there are *other* cognitive ability factors as well as non-ability factors which contribute to hypnotizability such as attitudes, motives, situational and interpersonal considerations, and perhaps certain personality characteristics.[5]

The present paper is deliberately restricted in scope, however, solely to these hypothesized ability components. A paper-and-pencil self-report questionnaire has been developed to elicit reports on experiences which may be indicative of these underlying abilities. The items of the questionnaire were gathered and refined over a period of several years.

The questionnaire is an attempt to devise a measure *independent of hypnotizability* of ability factors which are hypothesized as components of hypnotizability. The questionnaire is thus not restricted to items which appear to reflect the ability to suspend the usual generalized reality-orientation (tranceability) but includes also items which seem to pertain to *any* ability for "hypnotic-like" experiences occurring naturally in the normal course of living, independent of the use of special techniques (such as hypnosis, sensory-deprivation, drugs, etc.). It should thus be clear that these hypothesized abilities are of interest in their own right beyond their relationship to hypnotizability; they may provide an independent linkage to a variety of special states where alterations in consciousness have been descriptively noted.[6]

In an earlier publication normative data was presented on a briefer, less heterogeneous version of the present questionnaire (Shor, 1960). It was found that naturally occurring "hypnotic-like" experiences were fairly common in the normal college population and that the relative commonness of these experiences was consistent for two somewhat divergent strata of this population.

---

[4] A number of experimental studies have parallels to these views. See for example Barry, Mackinnon and Murray (1931); White (1937); Sutcliffe (1958).

[5] Research is in progress where these many factors are studied simultaneously.

[6] For example: road hypnosis, the break-off phenomenon, trance in other cultures, brain-washing, hallucinogenic drug experiences, sensory-deprivation, intense absorption and fascination, "mystic" experiences, peak experiences, and the inspirational phase of creativity.

Recently a number of other investigators have been studying the relationship between unusual personal experiences and hypnotizability. Aas (1960, 1961), Evans (1960), and London (1960) have been using, at least in part, some of our items with encouraging results.

B. THE PERSONAL EXPERIENCES QUESTIONNAIRE

The Personal Experiences Questionnaire in its original long form, as used in this research, is a paper-and-pencil self-report inventory of 149 items which inquires about a wide variety of naturally occurring "hypnotic-like" experiences. All questionnaire items were phrased positively so that the amount of agreement to each item was expected to correlate positively with criterion measures of hypnotizability.

On the front page of the questionnaire, space was allotted for the commonly requested identifying information such as subject's name, age, sex, and so forth. There followed a brief description of the questionnaire and instructions. This information is reproduced below.

> Description and Instructions. In this questionnaire we are interested in a variety of experiences which occur in the natural course of living and which tend frequently to be ignored. The questions are concerned only with experiences which occur naturally, not in dreams nor as the result of drugs nor of special techniques (such as hypnosis or sensory-deprivation).
>
> Normal individuals are able to recall a wide variety of such experiences when asked to do so. Please make an honest effort to remember as accurately as possible.
>
> You are to describe your experiences in two different ways. First in terms of frequency—that is, how often have you had the experience described. Secondly in terms of intensity—that is, how vivid and profound the single most intense experience was. *Intensity* is distinct from *frequency*. It is possible, for example, to have had a certain experience very frequently, but never to have had it particularly vividly or intensely. It is also possible to have had a certain experience rarely, perhaps only once in a lifetime, and yet with extraordinary vividness and intensity.
>
> You are to give the two kinds of answers in the following way:
>
> (1). *Frequency:* Under each question there is a scale with seven subdivisions: never, very rarely, rarely, occasionally, often, very often, and always. You are to read through each question and then rate yourself on this seven-point scale by placing a circle around the most appropriate answer.
>
> (2). *Intensity:* When referring to intensity forget about how frequently you have had the experience. Select only *the most intense experience* you have had. Then do one of three things:
>
> (*a*). if the single most intense of these experiences was quite vivid

> or profound place a single check mark in the right hand margin beside the question. (√)
>
> (*b*). if the single most intense of these experiences was *extremely* profound, intense or vivid place two check marks in the right hand margin beside the question. (√√)
>
> (*c*). if the single most intense of these experiences was not profound or vivid put nothing at all in the right hand margin.
>
> Check or double check for intensity as many or as few of the questions as you wish. However, answer *all* questions as to *frequency*. If in doubt make your best guess.

Among the topic areas covered in the questionnaire were the following overlapping categories: (*a*) the activities of sleeping or preparing for sleep; (*b*) the characteristics of dreams or dreaming; (*c*) the ability to concentrate and to achieve a quieting of mental activities; (*d*) the amount of freedom which the individual allows himself in emotional expression; (*e*) naturally occurring trance experiences; (*f*) suggestibility in everyday life situations; (*g*) temporary loss of awareness of generalized reality when mentally absorbed or preoccupied; (*h*) the extent of deliberate control of attentional processes; (*i*) sensory illusions in waking and hypnogogic states; (*j*) amnesias, absentmindedness, or unusual forgetting in everyday life; (*k*) naturally occurring illusions and hallucinations; (*l*) the extent to which an imaged or pseudo-reality has ever become temporarily indistinguishable from the actual reality; (*m*) artistic, mystical, transcendental, and profound religious experiences with altered consciousness; (*n*) detachment, depersonalization, and estrangement from self or reality.

The following are typical items in the questionnaire: (*a*) Have you ever had part of your body look strange and not part of your body at all? (*b*) Have you ever found yourself staring at something and for the moment forgotten where you were? (*c*) Do you ever get seasick at ocean movies? (*d*) Have you ever had the experience of walking in your sleep? (*e*) Have you ever felt drunk while sober? (*f*) Have you ever become so absorbed in listening to music that you almost forgot where you were? (*g*) Have you ever felt a oneness with the universe, a melting into the universe, or a sinking into eternity? (*h*) Have you ever been able to ignore pain? (*i*) Do you ever get unusually sleepy when reading dull material? (*j*) Has everything in your line of vision become blurry or strange, as if you were dreaming? (*k*) Do you find yourself unwittingly adopting the mannerisms of other people? (*l*) Have you ever had the experience of seeming to watch yourself from a distance as if in a dream?

The Personal Experiences Questionnaire was administered concurrently to two independent groups of subjects among batteries of other psychological tests, including measures of hypnotizability. The questionnaire data from one of these samples were analyzed first and without reference to the data from the other sample. The sample analyzed first will be called in this paper *the validation sample*. The results of analyses upon the validation sample compared against a scale of hypnotic susceptibility were then used as the basis for analyzing the questionnaire data in the remaining sample. This sample, analyzed second, will be called in this paper *the cross-validation sample*.

C. The Validation Sample

The section leader of an introductory psychology class of 51 students at Brandeis University administered to the group the Personal Experiences Questionnaire among a number of other paper-and-pencil questionnaires. With variable amounts of urging 29 of these students, or 57 per cent, later agreed to participate in hypnotic testing with the Stanford Hypnotic Susceptibility Scale (Weitzenhoffer and Hilgard, 1959). Nineteen of these students were male, ten were female.[7]

D. Reduction of Data

Item-validity coefficients of correlation were computed for each item of the Personal Experiences Questionnaire using as criterion the Stanford Hypnotic Susceptibility Scale scores. The main objective of this validating procedure was to select for cross-validation those items which seemed the most promising predictors of hypnotizability. For computational ease the simple phi coefficient (chi-square) was chosen as the correlation index. Statistical significance was determined in terms of the upper tail of the probability distribution. It was decided to select for cross-validation only those items which would yield a degree of positive association statistically significant at or beyond the .05 point; i.e., with a constant $N = 29$, a phi coefficient of $+ .31$ or above.

These item-validity phi coefficients were computed separately for the *frequency* and *intensity* questionnaire measures.

E. Frequency Measures

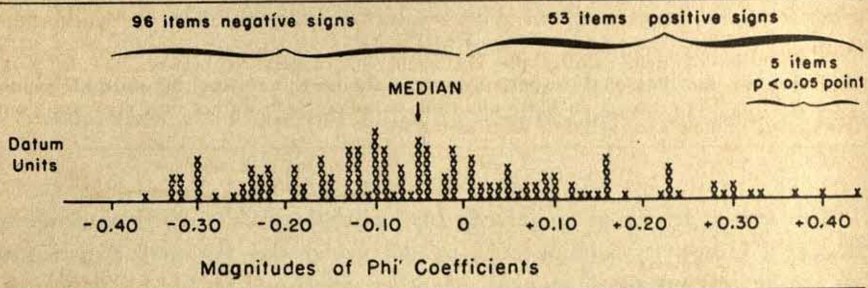
In Table 1 is presented the distribution of the item-validity phi coefficients for each item's *frequency* measure.

---

[7] The writers wish to thank Dr. Ricardo B. Morant of Brandeis University for making his introductory psychology class available to us. We also acknowledge the considerable help of Mr. John Stevens, section leader of this class, for administering the questionnaires and carrying out the hypnotic testing.

TABLE 1

Distribution of item-validity phi coefficients of correlation between the *Frequency* measures of the Personal Experiences Questionnaire and the Stanford Hypnotic Susceptibility Scale Scores in the Validation Sample. Each datum unit represents one questionnaire item's correlation with criterion.



1. *Findings*

Of the 149 items only 53 or 36 per cent yield the expected positive correlations with the criterion measure of hypnotic susceptibility. The median phi coefficient is —.05. The trend in this sample is thus in the direction opposite to prediction. Moreover, on the Null Hypothesis of equal binomial probability of positive and negative signs, chi-square is equal to 11.84 which with one degree of freedom has a corresponding two-tailed probability less than the .001 level which would be considered statistically highly significant had there been no previous prediction of direction. Only five items or 3½ per cent meet or exceed the minimum cut-off standard of phi = +.31, too few to be significant by combined probabilities, and certainly far too weak a demonstration of positive association to be of any practical use as a predictor of hypnotizability. Thus, except for the suspicion of a very slight and unexpected negative relationship, findings with the *frequency* measure are null.

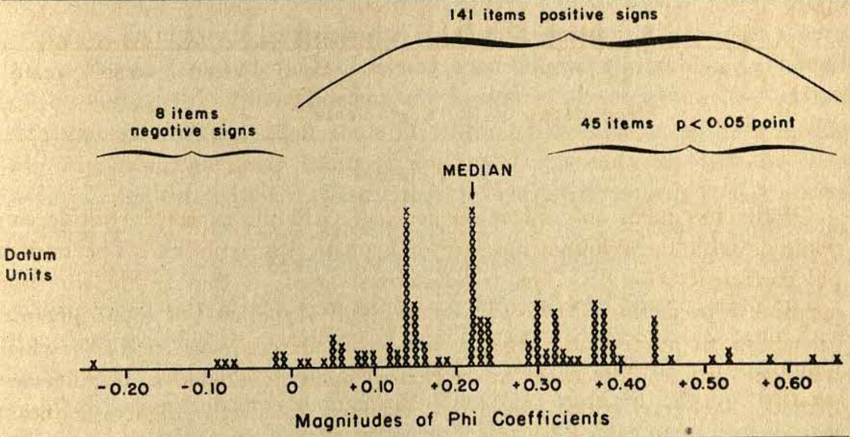
F. Intensity Measure

In Table 2 is presented the distribution of the item-validity phi coefficients for each item's *intensity* measure.

Of the 149 items 141 or 95 per cent yield the expected positive correlations with the criterion measure of hypnotic susceptibility. The median phi coefficient is +.22. Only eight phi coefficients have negative signs and all but one of these is a correlation lower than —.10. On the Null Hypothesis of equal binomial probability of positive and negative signs chi-square is equal to 117.2 with one degree of freedom. The exact probability of this occurrence was computed through the binomial expansion and the odds against this difference being a chance event are so astronomically huge ($p < 10^{-33}$) that there can be no reasonable statistical doubt of the existence of a positive association between the intensity measure and the validity criterion of hypnotic susceptibility. Moreover, since the datum units are themselves correlation coefficients the findings are very convincing statistical testimony to the consistency (reliability) among items. Forty-five items or 30 per cent of the total meet or exceed the minimum cut-off standards of phi = +.31. These statistically most promising 45 items scored for intensity were then selected for cross-validation.[8]

TABLE 2

Distribution of item-validity phi coefficients of correlation between the *Intensity* measures of the Personal Experiences Questionnaire and the Stanford Hypnotic Susceptibility Scale Scores in the Validation Sample. Each datum unit represents one questionnaire item's correlation with criterion.



G. The Cross-Validation Sample

The Personal Experiences Questionnaire was administered among a battery of other psychological tests to a carefully selected group of 25 college students from various universities in the Boston area. All but 6 of those selected were individuals who had had hypnotic training and had participated in hypnotic experimentation with our research project. In many cases their contact with us as hypnotic subjects had extended over a period of a year or longer.

---

[8] A copy of the 45 items as now in use with somewhat simplified scoring will be made available upon request. A complete listing of the 149 items with their respective *frequency* and *intensity* phi coefficients has been deposited with the American Documentation Institute and may serve as an item pool. Order Document No. 6847, from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

The purpose was to secure a sample consisting mainly of subjects whose level of hypnotic performance was already well-known, who then in repeated hypnotic sessions could be trained to their furthest limits of hypnotizability or at least to the point where it seemed most unlikely that any appreciable improvement would occur. Had a more random sampling procedure been used (as in the Validation Sample) a more normal distribution of hypnotizability undoubtedly would have emerged. Our purpose, however, was not to try to match the shape of the hypnotizability distribution in the college student population but rather to study more fully the two extreme ends of that distribution. The more stratified sampling procedure used resulted in a greater frequency of representation in the theoretically more interesting extremes.

H. Order of Procedures

The order of the experimental procedures reported in this paper utilized with these 25 subjects was as follows: (*a*) selection of subjects; (*b*) self-administration of the Personal Experiences Questionnaire; (*c*) administration of the Stanford Hypnotic Susceptibility Scale; and (*d*) diagnosis of maximum achieved depth in hypnotic evaluative sessions.

1. *The Personal Experiences Questionnaire*

The intensity scores for the 45 Personal Experiences Questionnaire items previously validated were summated into a single intensity measure for each of the 25 subjects. The odd-even reliability coefficient computed for these intensity measures, corrected by the Spearman-Brown formula, is $r = +.95$.[9]

2. *Stanford Hypnotic Susceptibility Scale*

The Stanford Hypnotic Susceptibility Scale was administered throughout by one of the writers (D. N. O'C.).

3. *Hypnotic Evaluative Sessions*

The controversy regarding universal hypnotizability remains unresolved; i.e., whether with unlimited time and ingenuity everyone eventually could be profoundly hypnotized. Nevertheless, our recurring experience has been that for most subjects most of the time a plateau in hypnotic performance is soon reached after which no appreciable improvement occurs regardless of the hypnotist, methods used, or amount of further training. It was thus decided for the purposes of this sample to define hypnotizability as the maximum hypnotic depth actually achieved in as many hypnotic training sessions as the experimenter needed in order to feel confident that a stable plateau had been reached.

---

[9] **The counterpart odd-even coefficient computed on the 45 validated items in the Validation Sample is $r = +.96$.**

Thus, regardless of any prior knowledge of a subject's level of hypnotic performance, a minimum of one further session was given. In these sessions the experimenter was allowed freedom to utilize any hypnotic techniques which seemed appropriate and to explore clinically any issues which might then help maximize performance. All hypnotic sessions were administered by one of the writers (M. T. O.). For estimates of inter-judge reliability another of the writers (R. E. S.) observed all of these sessions through a one-way screen with audio arrangements.

Both the experimenter and the observer independently formed ratings of maximal hypnotic depth achieved. These ratings may best be described as clinical diagnoses by experienced hypnotists based upon both objective hypnotic behavior and subjects' report. Four general groupings or categories of depth (hypnotizability) were established: (*a*) Less Than Light; (*b*) Light; (*c*) Medium; and (*d*) Deep. Plus and minus intermediate ratings were also assigned, but it was decided to ignore these lesser subdivisions for many of the first stages of data analysis. The experimenter's depth ratings were correlated with the observer's depth ratings and found to be just about identical ($r = +.96$). It thus seemed legitimate to average the two sets of ratings to form a single Over-all Depth Rating. The four major depth categories used are for all practical purposes entirely consistent with the understandings of these terms in common usage and descriptively correspond to the major divisions of the Davis-Husband scale (1931).

In Table 3 are presented the frequencies and percentages of subjects classified into each of the four major categories of hypnotic depth by the averaged experimenter-observer ratings. Also presented for each category are frequencies of new and previously evaluated subjects.

It may be observed that a roughly equal percentage of subjects were classified into each of the four categories of depth; i.e., the distribution is roughly rectangular. It is also of interest to note that all of the eight Deep Hypnosis subjects were individuals who had had prior evaluations.

TABLE 3
Frequencies and percentages of subjects in the four categories of hypnotic depth. Also presented are frequencies of new and previously evaluated subjects.

	OVERALL DEPTH RATINGS				
	Less Than Light	Light	Medium	Deep	Total
Frequency	7	6	4	8	25
Percentage	28	24	16	32	100
Had Prior Evaluations	6	4	1	8	19
New Subjects	1	2	3	0	6

I. Results

In Table 4 is presented a matrix of Pearson *r* coefficients of correlation with associated probabilities for the questionnaire and the two hypnotic depth measures. The variable hypnotizability is assumed throughout to have underlying continuity and thus the Pearson *r* coefficients are corrected for

TABLE 4
Matrix of Pearson *r* coefficients of correlation and relevant sampling data for the questionnaire and the two hypnotic depth measures. Coefficients are corrected for course grouping where appropriate. Stanford Hypnotic Susceptibility Scale data are not available for two of the Deep Hypnosis subjects which accounts for the decreased N.

	Personal Experiences Questionnaire Sum Intensity Checks	Stanford Hypnotic Susceptibility Scale
Overall Depth Rating	$r = +0.44^{*}$ $N = 25$	$r = +0.83^{**}$ $N = 23$
Stanford Hypnotic Suscept. Scale	$r = +0.46^{*}$ $N = 23$	

$^{*}p < 0.01$ point $\quad$ $^{**}p < 0.001$ point course groupings where appropriate as discussed by Guilford (1950; pp. 359-361).[10]

There is a high correlation between the two hypnosis measures, +.83, which means that about 70 per cent of the variance between these two quite different measures of hypnotic depth is accountable in terms of their measurements in common.

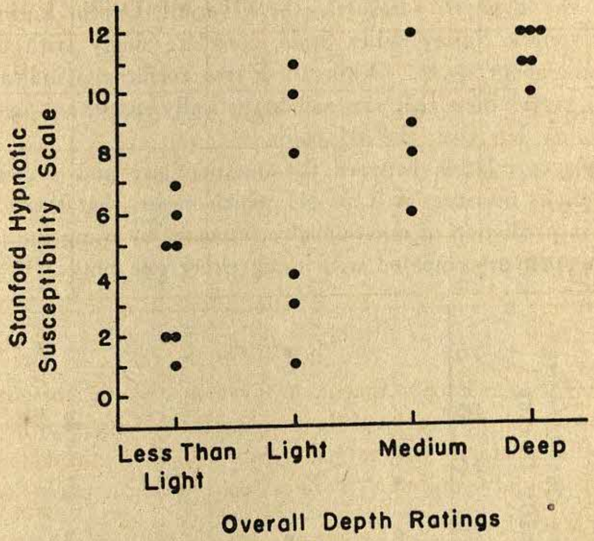


FIGURE 1
SCATTER-DIAGRAM OF THE CORRELATION BETWEEN THE STANFORD HYPNOTIC SUSCEPTIBILITY SCALE AND THE OVERALL DEPTH RATINGS

In Figure 1 is presented a scatter-diagram of this correlation. Note that all of the Deep Hypnosis subjects on whom Stanford Hypnotic Susceptibility Scale information is available received scores of 10 or more.[11] It is this tighter cluster of scores for the Deep Hypnosis subjects which has considerable influence in yielding a higher coefficient of correlation than might seem indicated by visual inspection of the scatter-diagram.

---

[10] Because of the questionable legitimacy of using parametric statistical procedures on some of our distributions Spearman rho coefficients of correlation, corrected for ties, were applied to our data as a conservative check on the corrected Pearson *rs*. In every instance the rank order coefficients were very closely comparable.

[11] The Stanford Hypnotic Susceptibility Scale is scored in terms of certain specific *objective* behavior. Inquiries with the Deep Hypnosis subjects after testing clarified that all failed items were *subjectively* successful; as, for example, the individual who objectively does not brush away a bothersome mosquito but who subjectively hallucinates doing so.

Observe also that 4 of the 7 subjects classified as Less Than Light in the Over-all Depth Ratings received Stanford Hypnotic Susceptibility Scale scores between 5 and 7, which is the middle region of the Stanford Scale's distribution. Notice also the wide distribution of Stanford Hypnotic Susceptibility Scale scores for the subjects classified into the Light Hypnosis category.

Both of the hypnosis measures, the Over-all Depth Ratings and the Stanford Hypnotic Susceptibility Scale, correlate singly with the questionnaire measure about +.45. While these two coefficients indicate only low to moderate correlations they are each statistically significant with an associated probability less than the .01 point.

A multiple correlation between the questionnaire and a combination of the two hypnosis measures is R = .47 which means that there is no appreciable gain in prediction of questionnaire responses by using the two hypnosis measures together as compared with using either one singly.

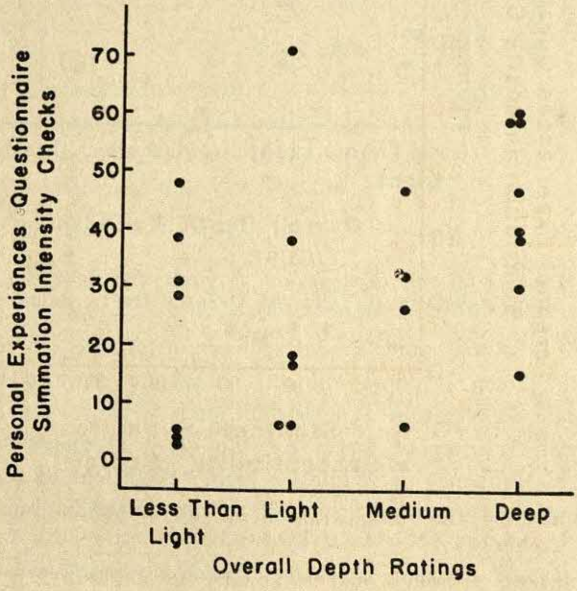


FIGURE 2
SCATTER-DIAGRAM OF THE CORRELATION BETWEEN THE QUESTIONNAIRE MEASURE AND THE OVERALL DEPTH RATINGS

In Figure 2 is presented a scatter-diagram of the relationship between the questionnaire measure and the Over-all Depth Ratings. By visual inspection of the scatter-diagram it is only with some effort that the correlation can be seen, and even then the association seems apparent only in the divergence between the Deep Hypnosis ratings and all three of the other ratings combined. The correlation between the questionnaire measure and the Over-all Depth Ratings *when the Deep Hypnosis ratings are excluded* is $r = +.17$; i.e., the relationship virtually disappears. In other words, the questionnaire measure does not appear appreciably related to this particular measure of hypnotizability from Less Than Light up through Medium Hypnosis.

In Figure 3 is presented a scatter-diagram of the relationship between the questionnaire measure and the Stanford Hypnotic Susceptibility Scale scores.

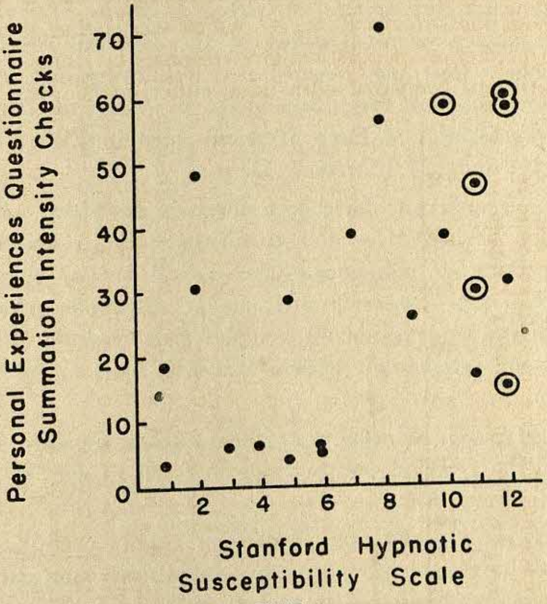


FIGURE 3
SCATTER-DIAGRAM OF THE CORRELATION BETWEEN THE QUESTIONNAIRE MEASURE AND THE STANFORD HYPNOTIC SUSCEPTIBILITY SCALE

The correlation, although mathematically no higher than the one just discussed, appears visually to be more apparent over a wider range of the hypnotizability measure. Encircled in the scatter-diagram are the datum units for the subjects classified as Deep in the Over-all Depth Ratings. Re-computing the correlation *when excluding these Deep Hypnosis subjects* yields a coefficient of $r = +.33$ which suggests some association but does not quite achieve statistical significance ($p > .10 < .20$ point).

J. Distinctions Within the Deep Hypnosis Category

It has long been the clinical impression of the writers that many of the distinctions on the continuum of hypnotizability of the greatest theoretical interest lie *within* the Deep Hypnosis region. It has also been our view that the psychological distance on the hypnotizability continuum occupied by the Deep Hypnosis region is even greater than the distance occupied by all of the categories of lesser hypnotic depth combined. In other words, we feel that the continuum of hypnotizability might be best depicted schematically in logarithmic form, as in Figure 4.[12]

One of the primary guiding objectives behind the design of our research was to explore the relationship between the Personal Experiences Questionnaire and hypnotizability *within the Deep Hypnosis region.* While only 8 of the 25 subjects were classified in the Over-all Depth Ratings as Deep, exceptional effort was devoted to carefully evaluating the level of hypnotic performance of these 8 subjects in order to sub-classify them into the three categories of Deep Hypnosis cited in Figure 4; (*a*) Deep, (*b*) Very Deep; and (*c*) Extremely Deep.

In order for a subject to have been classified anywhere within the Deep Hypnosis region he had to be able to manifest readily and reliably all of the following criterion phenomena when sampled: (*a*) suggested blanket amnesia which resists concerted attempts at disruption after awakening; (*b*) somnambulism and reasonably complex post-hypnotic effects; and (*c*) hallucinations and interactions with a fantasied environment (for example, age-regression).

The Deep Hypnosis subjects were then further classified into the three subdivisions of the Deep Hypnosis region in terms of very careful diagnoses of the following:

a. Deep. All of the criterion phenomena readily producible either very well or at least to a moderate extent but subjectively not always fully convincing to the subject himself as he reports afterwards on what it was like at the moment of the experience.

b. Very Deep. All of the criterion phenomena readily producible but occasionally in a less than fully complete form. Mostly subjectively convinc-

---

[12] The continuum of hypnotizability described here as unidimensional is no contradiction to the view that hypnotic depth may have underlying multidimensionality. In most hypnoses with conventional hypnotic techniques the effort is to intertwine all factors for mutual influence. Only when gross imbalances exist among these factors would a unidimensional continuum be misleading (Shor, 1962a; 1962b).

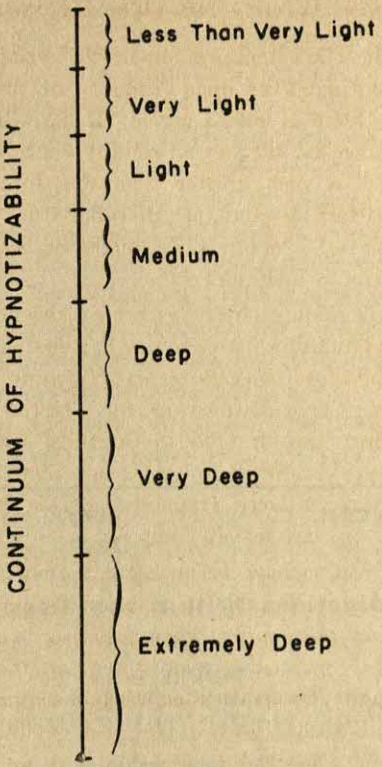


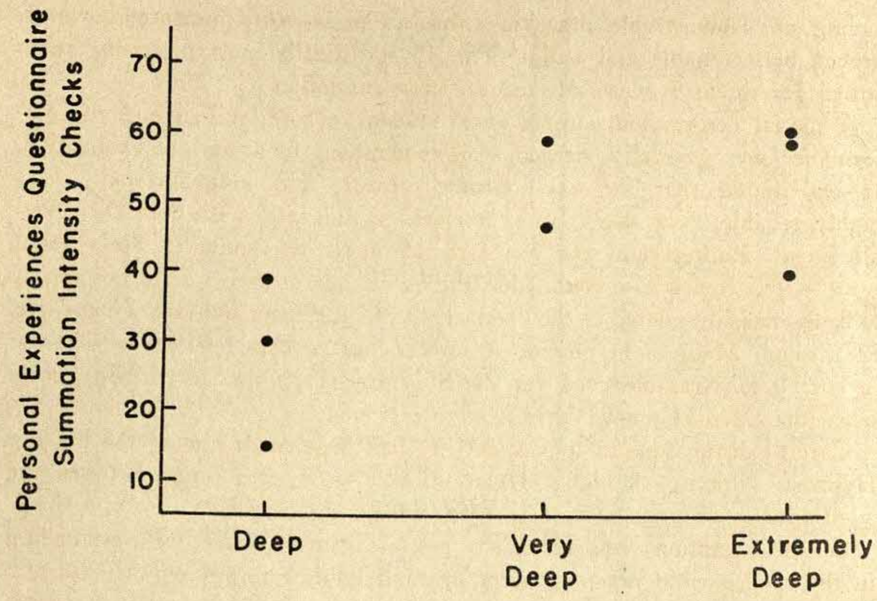
FIGURE 4
Schematic representation of the continuum of hypnotizability as conceived by the writers in terms of their clinical experience. Note that the categories drawn along the continuum signifying deeper hypnotic depth are assigned increasingly greater amounts of psychological distance, as represented in logarithmic form. Note also that the three categories of the Deep Hypnosis region occupy more than half of the total distance. The Figure represents a statement of impressions, of hypotheses, and has not been formally deduced from experimental evidence.

ing to the subject but occasionally with some uncertainty about its utter reality at the moment of the experience.

c. Extremely Deep. All of the criterion phenomena readily producible in their most complete form with the fullest subjective conviction.

K. RESULTS

In Figure 5 is presented a scatter-diagram of the correlation between the questionnaire measure and the distinctions within the Deep Hypnosis category.



Distinctions Within the Deep Hypnosis Category

FIGURE 5
SCATTER-DIAGRAM OF THE CORRELATION BETWEEN THE QUESTIONNAIRE MEASURE AND THE DISTINCTIONS WITHIN THE DEEP HYPNOSIS CATEGORY

There appears to be a marked relationship with no overlap between the questionnaire scores for Deep as compared with the remaining two subdivisions. The corrected Pearson r for this relationship is +.84 which is of course highly significant statistically ($p < .005$ point). This finding appears to indicate that the questionnaire measure has high predictive potency in the deeper region of the hypnotizability continuum. Moreover, the correlation is significantly greater statistically than the questionnaire measure's virtually inconsequential predictive potency in the Less Than Light through Medium Hypnosis region cited earlier.

L. RECAPITULATION OF THE EXPERIMENTAL RESULTS AND NON-INDEPENDENCE AS A POSSIBLE SOURCE OF CONFOUNDING

On a random sample of 29 subjects the Personal Experiences Questionnaire was validated as a predictor of hypnotizability as measured by the Stanford Hypnotic Susceptibility Scale. The *frequency* measure emerged as having no demonstrable practical value. The *intensity* measure, however, proved both reliable and valid. The 45 statistically most promising items scored for intensity were selected for cross-validation.

A special non-random sample of 25 students whose level of hypnotic performance was generally already well-known was used for cross-validation. It was found that the questionnaire measure was both internally very highly reliable, $r = +.95$, and correlated significantly with both the Overall Depth Ratings and the Stanford Hypnotic Susceptibility Scale about $r = +.45$. It was also found that the questionnaire measure did not appear to be appreciably related to the Over-all Depth Ratings from Less Than Light up through Medium Hypnosis. A similar, but perhaps less severe, decrease in correlation was observed for the Stanford Hypnotic Susceptibility Scale when the Deep Hypnosis subjects were excluded.

Careful distinctions in hypnotizability had been made among the 8 Deep Hypnosis subjects. A high correlation, $r = +.84$, was found between the questionnaire measure and these Deep region distinctions.

Scientific caution requires that a possible general source of confounding in the experimental design be kept in mind while interpreting the evidence. In planning the cross-validation sample it was decided deliberately to forego double-blind methodology. For this particular special sample it was not possible to achieve rigorous independence of judgments in regard to the characteristics of subjects already so well-known. It was felt, however, that additional purity at this stage could only have been had at too great a sacrifice in ground covered. It would thus be unwise to interpret our results as not requiring more scrupulous experimental corroboration.

In spite of reservations appropriate to the lack of independence it is warranted to conclude that the results are a *modestly convincing* demonstration that the questionnaire measure evolved is very highly reliable and is a valid predictor of hypnotizability, and most clearly so in the deeper region of the hypnotizability continuum.

M. DISCUSSION AND THEORETICAL EXTRAPOLATION

The over-all generalization which may be drawn from the study is as follows: there is reasonable evidence that the questionnaire as evolved predicts hypnotizability, especially in the deeper region of the hypnotizability continuum; it is less predictive, if at all, in the lighter region.

This generalization is fully congruent with our initial theoretical formulations provided one additional theoretical assumption be made: that just about everyone in the normal college population (from which all of our samples have been drawn) has sufficient of the hypothesized cognitive ability components to achieve an average amount of hypnotizability (Light to Medium depth) if other conditions are favorable.[13] From this assumption it follows that it is these other conditions (the hypothesized non-ability components, such as attitudes and motives) which are the primary determinants of hypnotizability for college student subjects who classify into the lighter and middle region of the hypnotizability continuum. It is only in the Deep and deepest region of the hypnotizability continuum that the possession of *exceptionally high* ability becomes an important determinant. Moreover, it is also in these deeper regions that the non-ability factors should in theory become less predictive since they would almost assuredly have to be no worse than slightly unfavorable or the subject would not have achieved Medium or greater depth in the first place. It thus necessarily follows that: (*a*) the hypothesized ability components are the primary determinants of hypnotizability in the deeper region of the hypnotizability continuum; and, contrariwise; (*b*) the hypothesized non-ability components are the primary determinants of hypnotizability in the lighter region of the hypnotizability continuum. It follows further that a multiple correlation composed of both ability and non-ability factors would predict hypnotizability along the entire continuum of hypnotizability.

These theoretical deductions, while fully consistent with our findings, are, of course, extrapolations beyond them. As a convenient aid to comprehending the many interrelationships of these theoretical extrapolations, a schematic representation of them is presented in graphic form in Figure 6.

In the Figure are represented the shapes of the theoretically expected correlations for college student subjects between depth on the continuum of hypnotizability and the two hypothesized major component factors. The continuum of hypnotizability, in logarithmic form, is presented along the abscissa. The Figure thus represents the superimposition of two 2-dimensional graphs sharing the common abscissa, and therefore is actually 3-dimensional.

The left hand ordinal scale represents six general qualitative distinctions in *amount of ability components;* none, very little, some, moderate, considerable, and very considerable. The general shape of this variable's correlation with hypnotizability is represented as the area enclosed by the solid outline on the graph.

---

[13] An assumption which is indirectly supported by our normative data cited earlier (Shor, 1960). In samples which include many individuals of lesser cognitive abilities than college students we would not expect this assumption to hold.

The right hand ordinal scale represents six general qualitative distinctions in extent favorableness of non-ability components; very unfavorable, unfavorable, slightly unfavorable, slightly favorable, favorable, and very favorable. The general shape of this variable's correlation with hypnotizability is represented as the area enclosed by the broken outline (dashes) on the graph.

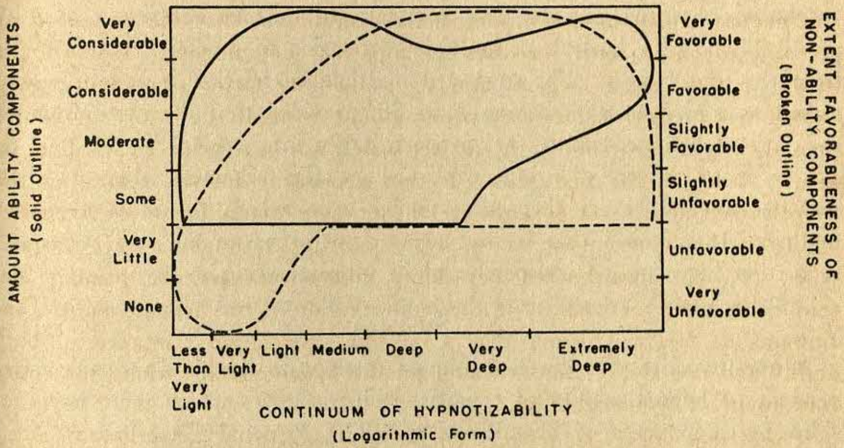


FIGURE 6

SCHEMATIC REPRESENTATION IN GRAPHIC FORM OF THE INTERRELATIONSHIPS OF THE THEORETICAL EXTRAPOLATIONS FOR COLLEGE STUDENT SUBJECTS FEATURES ARE DISCUSSED IN DETAIL IN THE TEXT

The area enclosed by the solid outline is depicted as predictive of hypnotizability mostly in the deeper regions; the area enclosed by the broken outline is depicted as predictive of hypnotizability mostly in the lighter regions.

The solid outline is horizontally truncated along the base to represent the assumption that just about everyone in the normal college population has at least a modest amount of the hypothesized ability components. The broken outline also is horizontally truncated along the base from about Medium hypnotizability and deeper. This is to represent the expectation that an individual who achieves Medium or deeper hypnotizability would almost assuredly have no worse than slightly unfavorable non-ability components else he would not have achieved Medium or deeper hypnotizability in the first place.

Observe that the vertically highest portion of the solid outline in the deepest region of hypnotizability is no higher than the vertically highest portion in the Light to Medium region. Note however that there is a dip in between.

This theoretically predicted dip is a logically necessary deduction from the above hypothesis that an individual who achieves Medium or deeper hypnotizability would almost assuredly have no worse than slightly unfavorable non-ability components. It follows that an individual with no worse than slightly unfavorably non-ability components *and* very considerable ability components would probably achieve hypnotizability in the *deepest* regions rather than just in the Deep.

Observe that the way the two correlational outlines intercross suggests the likelihood of a fairly high multiple correlation to predict hypnotizability from the two factors. The suggested multiple correlation is hopefully displayed as somewhat approaching $R = .80$, provided that the non-linearities and lack of homoscedasticity are properly taken into account.[14]

The many interrelationships expressed by this schematic diagram may serve as a fairly exact theoretical model upon which to compare future findings. It is hoped that it may prove fruitful as an aid and stimulant to further investigation irrespective of its ultimate accuracy.

N. SUMMARY

A questionnaire was developed in an attempt to devise a measure independent of hypnotizability of cognitive ability factors which were hypothesized as components of hypnotizability. The Personal Experiences Questionnaire was constructed as a paper-and-pencil self-report inventory of 149 items which inquired about a wide variety of naturally occurring "hypnotic-like" experiences. On a random sample of 29 college student subjects the Personal Experiences Questionnaire was validated as a predictor of hypnotizability as measured by the Stanford Hypnotic Susceptibility Scale. The 45 statistically most promising items were selected for cross-validation. A special non-random sample of 25 college student subjects whose level of hypnotic performance was generally already well-known was used for cross-validation. In addition careful distinctions in hypnotizability were made among the 8 most deeply hypnotizable subjects.

The results were a modestly convincing demonstration that the questionnaire as evolved is a predictor of hypnotizability, and most clearly so in the deepest region of the hypnotizability continuum. The findings are interpreted in terms of theoretical formulations where both ability factors and

---

[14] "Common experience shows that the validity coefficient for a single test may be expected within the range from .00 to .60, with most of them in the lower half of that range. Validity coefficients for composite scores based upon combinations of several different kinds of tests are likely to be distinctly higher, ranging up to .80 in rare instances but hardly ever above the latter figure." (Guilford, 1950; p. 165.)

non-ability factors (such as attitudes and motives) are viewed as components of achieved hypnotizability. Ramifications are discussed and theoretical extrapolations from the data are presented as an aid and stimulant to further investigation.

REFERENCES

1. AAS, A. Hypnotizability and non-hypnotic experiences, presented at the California Psychological Association Convention, 1960.
2. ————. Personal Communication, 1961.
3. BARRY, J., JR., MACKINNON, D. W., & MURRAY, H. A., JR. Studies in personality. A. Hypnotizability as a personality trait and its topological relations. *Human Biology,* 1931, **3**, 1-36.
4. DAVIS, L. W., & HUSBAND, R. W. A study of hypnotic susceptibility in relation to personality traits. *J. Abnorm. Soc. Psychol.,* 1931, **26**, 175-182.
5. EVANS, F. J. The prediction of hypnotic phenomena from questionnaire items. Unpublished manuscript, University of Sydney, Australia, 1960.
6. GILL, M. M., & BRENMAN, M. Hypnosis and Related States. New York: International Universities Press, 1959.
7. GUILFORD, J. P. Fundamental Statistics in Psychology and Education, second edition. New York: McGraw-Hill, 1950.
8. KUBIE, L. S., & MARGOLIN, S. The process of hypnotism and the nature of the hypnotic state. *Amer. J. Psychiat.,* 1944, **100**, 611-622.
9. LONDON, P. Personal Communication, 1960.
10. SHOR, R. E. Hypnosis and the concept of the generalized reality-orientation. *Amer. J. Psychother.,* 1959, **13**, 582-602.
11. ————. The frequency of naturally occurring "hypnotic-like" experiences in the normal college population. *Int. J. Clin. Exp. Hyp.,* 1960, **8**, 151-163.
12. ————. Three dimensions of hypnotic depth, to be published circa 1962a.
13. ————. A technical handbook for the systematic definition of the construct hypnosis, to be published circa 1962b.
14. SUTCLIFFE, J. P. Hypnotic behavior: fantasy or simulation? Doctoral dissertation, University of Sydney, Australia, 1958.
15. WEITZENHOFFER, A. M., & HILGARD, E. R. Stanford hypnotic susceptibility scale, Palo Alto: Consulting Psychologists Press, 1959.
16. WHITE, R. W. Prediction of hypnotic susceptibility from a knowledge of subjects' attitudes. *J. Psychol.,* 1937, **3**, 265-277.
17. ————. A preface to a theory of hypnotism. *J. Abnorm. Soc. Psychol.,* 1941, **36**, 477-505.

Studies in Hypnosis Project
Harvard Medical School
74 Fenwood Road
Boston 15, Massachusetts

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 77-80.

CONTROLLED ASSOCIATION TIME AS A FUNCTION OF WORD FREQUENCY*[1]

Department of Psychology, The Pennsylvania State University

John F. Hall

A. Problem

In a previous study, Hall and Ugelow (1957) reported that response latencies in a free association situation were related to the frequency of the presented stimulus word. Briefly, their results indicated that such latencies were significantly longer to low frequency words than to high. An analysis of the number of times a subject blocked to the high and low frequency words provided confirming evidence for the finding, with subjects blocking significantly more frequently to the low frequency words than to high. Finally, an examination of the response words revealed that there was greater communality (i.e., smaller range) of responses to the high frequency words than to the low.

The purpose of the present study was an attempt to determine if the frequency of the stimulus word played a similar role in the controlled association situation. More specifically, we "controlled" the subject's response by asking him to respond with a word which, together with the stimulus word, made up a meaningful combination of words.

B. Method

1. *Subjects*

Thirty-four students (24 males and 10 females) were recruited from an introductory psychology class at The Pennsylvania State University.

2. *Materials*

The construction of the word list was as follows. Fifty-six six letter adjectives were selected from the Thorndike-Lorge *Teacher's Word Book of 30,000 Words* (1944). Twenty-eight had a low frequency count (1 to 5 occurrences per million words) while the other 28 had a high frequency (50 or more occurrences per million words). The words used were selected

---

* Received in the Editorial Office on September 5, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This research was supported in part by The National Institute of Mental Health.

as follows. A table of random numbers was consulted to provide a list of page numbers ranging from 1 to 208 (the number of pages in the book). Turning to the page indicated by the first random number, the first six letter adjective which had the required frequency was chosen. If the page did not provide an appropriate word, succeeding pages in the Word Book were examined until such a word was found. This process was repeated until the 56 words to be used in the experiment were obtained.

3. *Apparatus*

The words were presented with a Revere 888 automatic slide projector. It was modified so that the depression of an experimenter's key projected the slide and also started a Standard Electric Timer. When *S* responded verbally, he lifted a telegraph key which stopped the timer and provided a latency measure.

4. *Procedure*

After the *S* had reported to the experimental room, he was comfortably seated in a chair to which was attached the telegraph key. *E* read a short paragraph indicating the nature of the task. *S* was instructed to depress the key when given a ready signal by *E*, and to lift his finger when he responded. Five to 10 stimulus words were presented to better acquaint the *S* with the task as well as to give him some practice in lifting the key when he made the verbal response. The 56 words were then presented, with the *S*'s responses and response times being recorded.

C. RESULTS AND DISCUSSION

Of 1904 possible responses from the 34 subjects, 49 responses were lost because of (*a*) mechanical or subject difficulty, or (*b*) the subject failed to provide a response which, in the best judgment of *E*, did not represent a meaningful combination of words. Analysis of these responses indicated that 23 of them were to high frequency stimulus words and 26 to low.

A further 168 responses were lost because *S* blocked to the stimulus word, and to avoid embarrassing *S* by waiting too long while he groped for a response, *E* stopped the clock and went on to the next word. Latencies from such responses were not included in the latency analysis but have been examined separately. The latency analysis then, was based upon a total of 1697 responses.

The statistical analysis was based on the mean latencies to the high and low frequency words obtained separately for each *S* and then averaged. The mean latencies, in hundredths of seconds, were:

High frequency words	465.1
Low frequency words	492.7

The difference, using the difference method, was quite significant ($t = 3.15$, $p < .01$). Only six of the 34 *S*s provided a more rapid mean association time to the low frequency words than to the high.

The distribution of blocking provided additional evidence for the finding that long association times were associated with low frequency words. The mean number of blocks for all *S*s to high and low frequency words was:

High frequency words	.91
Low frequency words	5.79

This difference, using the difference method, is again highly significant ($t = 4.78$, $p = < .01$). Only a single *S* blocked more to high frequency words than to low.

A second aspect of this investigation was to examine the communality of responses provided to both the high and low frequency words. The early study of Thumb and Marbe (quoted by Woodworth and Schlosberg, 1954) as well as the more recent one by Hall and Ugelow (1957) revealed that high frequency stimulus words provided greater communality of response than did low frequency words. The number of different responses to each stimulus word was obtained and divided by the total number of responses to that word. (In determining the number of different responses, a response which was plural in form was considered to be the same as the singular). The lower the resulting value, the higher the communality of the responses; that is, the more they are concentrated among relatively few response words. The mean values for the high and low frequency stimulus words were as follows:

High frequency words	.515
Low frequency words	.549

Although there appears to be a slight trend for the high frequency words to have greater communality than low, the difference is not significant.

These results confirm the general findings obtained when the free association situation is used. Thus, when a subject is asked to provide either a "free" or "restricted" response to a stimulus word, the frequency of the stimulus word plays an important role in determining his speed of responding. With regard to the communality of response, no difference was obtained between high and low frequency stimulus words when the controlled association situation was used. This is in contrast to the findings obtained with free association. A comparison of the communality values obtained in this study with those obtained in the previous one suggests that the controlled situation reduces somewhat the range of responses made to the low frequency words, while at the same time, increases the number of responses made to high frequency words.

D. Summary

This study was designed to determine the relationship between the frequency of the stimulus word, as indicated by the Thorndike-Lorge frequency count, and controlled association time. The controlled association situation was one of requiring the subject to provide a response which, together with the stimulus word, made a meaningful combination of words. Analysis of the association times to 56 words for 34 subjects revealed that latency was significantly shorter to high frequency words than to low. An analysis of the blocks made to the high and low frequency stimulus words provided confirming evidence. In contrast to the free association situation, however, an examination of the response words did not reveal any relationship between frequency and communality.

References

1. Hall, J. F., & Ugelow, A. Free association time as a function of word frequency. *Can. J. Psychol.,* 1957, **11**, 29-32.
2. Thorndike, E. L., & Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Teacher Coll., Columbia Univ., Bureau of Publications, 1944.
3. Woodworth, R. S., & Schlosberg, H. Experimental Psychology. New York: Henry Holt, 1954.

Department of Psychology
Pennsylvania State University
University Park, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 81-90.

A THEORY OF EGO PATHOLOGY AND EGO THERAPY*

Western Psychological Center, Sherman Oaks, California

STEWART B. SHAPIRO

A. INTRODUCTION

The purpose of this paper is to offer the beginnings of a theory of ego pathology and ego therapy which may help to clarify the current conflict between certain "structured" and "unstructured" positions in a theory of psychotherapeutic technique.

B. CRITIQUE OF THE PSYCHOANALYTIC (STRUCTURED) POSITION

Psychoanalytic literature is currently filled with so-called parameters—new techniques designed to treat the wide variety of patients needing help but not amenable to more or less classical psychoanalysis. The introduction of these many deviations from the "classical model" suggests that psychoanalytically oriented therapists as well as others, are no longer satisfied with the therapeutic results of traditional psychoanalytic methods.

The above may be true not only because psychoanalysis, as a therapy, was never intended to apply to a broad range of patients, (Waelder 14), but also because even applied to the problems of neurosis it leaves considerable to be desired as a therapy.

The fact that psychoanalysis or any other form of psychotherapy awaits rigorous scientific validation is only part of the problem. Another difficulty which haunts all therapies is the tendency for technique to run far ahead of the theory of technique. This is true in psychoanalysis despite the efforts of Freud (3), Fenichel (2), Menninger (10), Hartmann (6), Kris (8), and Loewenstein (9) and many others. As Menninger (10) has pointed out, most work has been done on theory or technique, and very little on the theory of technique. The works of Rogers (12), Pascal (11), Barker and Sundland (1), Greenspoon (5) and Shaw (13) have been notable exceptions to this from various non-psychoanalytic positions.

C. CRITIQUE OF THE "UNSTRUCTURED POSITION"

There is a strong contemporary trend in the field which deserves mention at this point. From the most widely divergent sources one hears a "theory"

---

* Received in the Editorial Office on September 5, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

of therapeutic technique in the form of the following injunction to the therapist: "Be yourself; be honest; don't rely on technique; use your own feelings (countertransference); especially, don't impose a façade, attitude or technique on yourself if it doesn't fit your personality; throw away the insipid non-directive smile or the impassive blank screen analyst role; do what comes naturally!"

In a preliminary publication of their study on the methods and attitudes of psychotherapists, Barker and Sundland (1) reported one general factor which cuts across sixteen dimensions by which therapists may be reliably characterized. This factor, which can be considered a continuum, was by far the most significant single dimension on which to compare therapists. Various labels have been attached to this continuum by these investigators and their colleagues as follows: Objective vs. Subjective, Cerebral vs. Visceral, Impersonal vs. Personal, Analytic vs. Existential and Observer vs. Participant. The "be yourself" injunction to the therapist clearly falls near one pole of this continuum which the writer prefers to call the "unstructured position" in psychotherapy as opposed to the "structured position" which emphasizes conceptualization, training, planning and interpretation of the unconscious.

It seems to this writer that there is at least one major theoretical difficulty with the unstructured position, and that is; "When you *are* yourself, what self are you? Are you the same self that would come out anyway, with or without technique? Are you a self, identified with the impulses of your patient, or with the patient's defenses, or with his super-ego?" One might answer this by saying that he is identified with the "constructive forces" in the patient's ego. His strong ego identifies with and reinforces the patient's weak ego. This "answer" like so many in this area has to await verification, but it is the writer's hypothesis that it is far less correct than many structured *and* unstructured therapists would like it to be.

The writer takes the position that there is no such thing as a single psychological self. There are only subselves more or less tied together and unified in an identity, depending on the relative strength of the synthetic function of the ego. It may be all right to "be yourself" as a therapist if you have some idea of *which* self you are being *when* and for what purpose, and with which of the patient's selves you align your various selves. It may be more important to know whether you are a consistent or inconsistent reinforcer for your patient's subselves or whether you *model* for your patient uncontrived self-disclosure in a fluid, well-communicating self-system than to "be yourself."

D. Modification of the Psychoanalytic Position

The theoretical position taken in this paper is essentially a modification of Freud's (4) structural hypothesis and the ego psychology of Hartmann (7). The following diagram (Figure 1) schematically illustrates the personality from this frame of reference before modification by the writer.

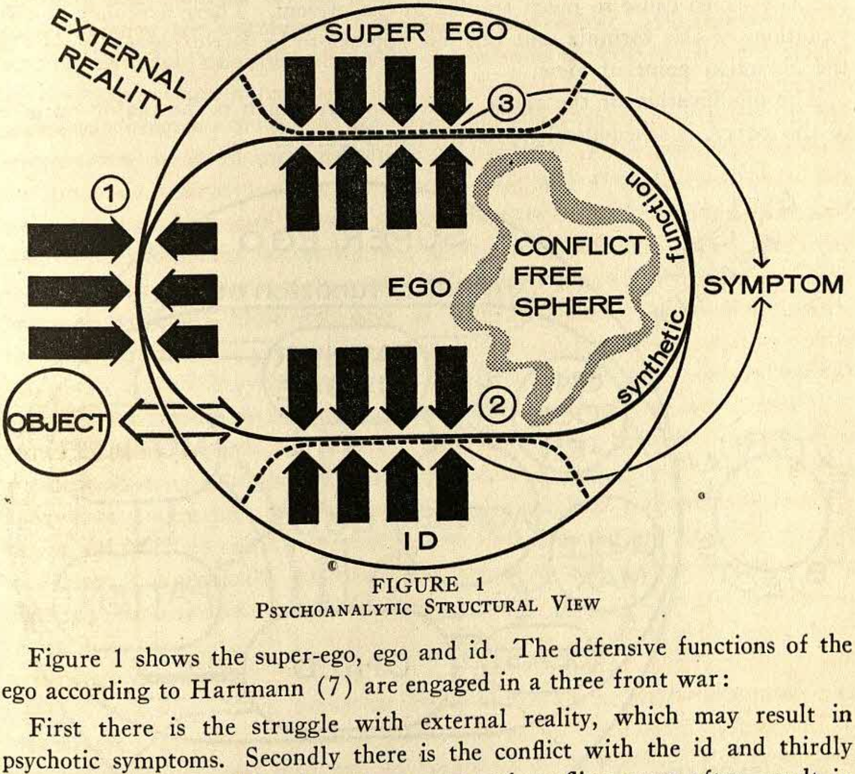


FIGURE 1
Psychoanalytic Structural View

Figure 1 shows the super-ego, ego and id. The defensive functions of the ego according to Hartmann (7) are engaged in a three front war:

First there is the struggle with external reality, which may result in psychotic symptoms. Secondly there is the conflict with the id and thirdly with the super-ego. The latter two types of conflicts most often result in neurotic symptoms. However, since 1939, ego psychology has been concerned with less defense-oriented concepts, such as the conflict-free spheres of the ego, the role of the ego in object relations, primary and secondary autonomy of the ego, and its synthetic function which ties the whole personality together. Such concepts as neutralization of energy and regression in the service of the ego are omitted from Figure 1 only in the interest of simplicity.

According to Menninger (10), one proceeds in psychoanalysis from reality conflicts and current object relations to the analyst (through the transference), and then through the regression to the past conflicts and past object relations. The path is from reality to transference to the past. Full interpretation by the analyst involves making the patient consciously aware of the remarkable commonality of current reality behavior, behavior in the transference and behavior with the significant objects of the past. This is supposed to undo the unwarranted generalizations and distortions from the past which cause so much trouble in the present. There are many slight variations of this formula, but this one appears to be fairly representative of the Freudian point of view.

The modification of the above point of view, which is the major purpose of this paper, is schematically presented in Figure 2.

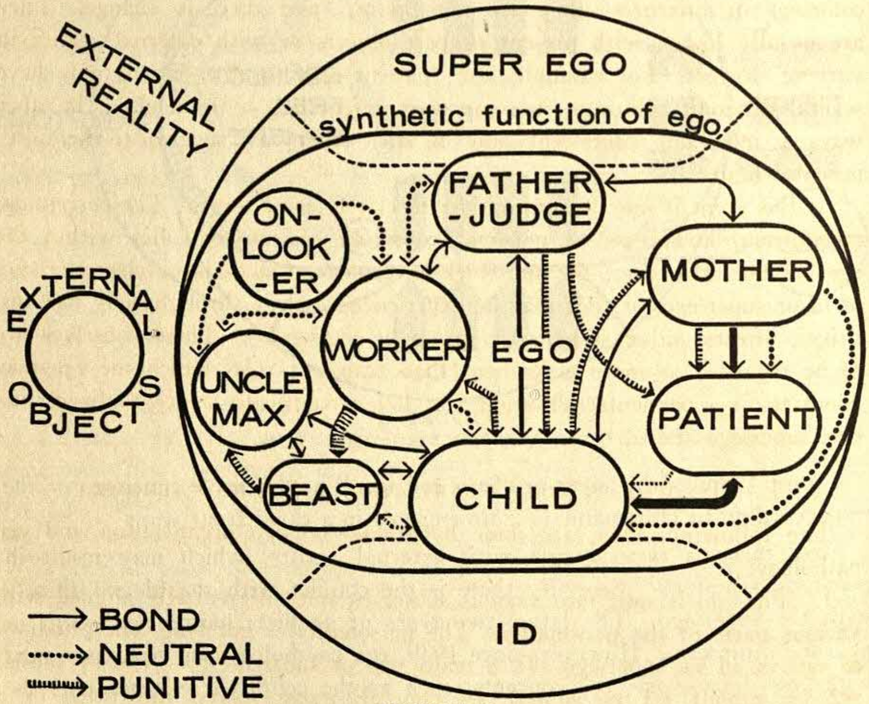


FIGURE 2
REVISED SUB-SELF VIEW

Figure 2 presents the personality from the "subself" point of view. As in the psychoanalytic system, there is a super-ego, an ego and an id. The major change, however, is within the ego, itself. According to this view the ego is made up of subselves. For purposes of illustration Figure 2 delineates eight subselves. There are the "mother" and "father or judge" subselves close to the super-ego. The "onlooker" and "Uncle Max" subselves are closer to external objects and external reality, and the "patient" and "worker" subselves occupy intermediate positions. The "child" and "beast" subselves are closer to the id.

How well the total personality functions depends on the relationship between these various subselves and the strength of the synthetic function of the ego which coordinates the entire organization. For example, there may be punitive, neutral or affectional relationships between these subselves as in any family group.

These subselves are on different levels of consciousness. They have different positions of dominance within the self system. They also have different colorings or mixtures—they are not "pure," and they do change. They are usually linked with present or past objects or with external reality in varying degrees. For example, the "patient self" in the patient may have several different colorings, in some ways resembling a timid child, in other ways a rebellious adolescent and in still others a "superior-to-therapist" parental figure.

At this point it may be appropriate to clarify some terms. The perception, recognition, awareness or personal construct of subselves lies within the ego system. However, the subselves have parts, functions or characteristics of id or super-ego, or they may be experienced mainly through body feelings, drives, affects, value systems or group belongingness. These subselves are to be regarded as more permanent than transient roles which the ego may adopt to fit a particular situation, but less categorically conceptualized than ego, super-ego and id.

E. HYPOTHESES EGO ORGANIZATION AND EGO PATHOLOGY

The following, then, are some hypotheses on ego organization and ego pathology.

1. The ego is split into various subselves and differentially aligned with various parts of the personality. The personality is not best conceptualized as ego vs. id vs. super-ego. It is really ego or subself #1 vs. ego or subself #2 vs. subself #3 vs. subself #4 etc., all more or less integrated by a superordinate identity corresponding to the synthetic function of the ego.

2. The major pathology of an adult personality "in trouble" is not inter-institutional or interpersonal, it is intra-institutional. The main source of current difficulty is not the symptom production or wishes vs. defenses or in pathological interpersonal relations. It is within the ego organization

itself . . . a house divided against itself, a kind of civil war between the subselves.

For example (see Figure 2), subself #1 may be primarily aligned with the conscience or the original mother as a scolding, critical self, but a self which also makes some objective observations and may even express concern or affection for the other selves. Subself #2 may be aligned primarily with the id . . . a childhood self age eight or nine. This child-self may be negative, angry and hurt. It may want to "move out of the house" (i.e., out of the ego, away from the critical subself #1), but it may be too timid to do so. Subself #3 may be aligned with a benevolent parent surrogate (e.g., "Uncle Max"). This self is more reasonable; it may take #2's side in a fight, but generally it is closer to the conflict-free spheres of the ego. This self is able to see and state #2's purpose and value to the total organization. Ego psychologists would say that subself #3 operates on more neutralized energy than #1 and #2, and that it is closer to autonomy from the drives.

3. More specifically the ego pathology is due to: (*a*) Poor communication and tensions among the subselves because of internalization of real object tensions (e.g., parents vs. child, mother vs. father, sibling vs. patient, value vs. value, society vs. individual. (*b*) The wish for one or more subselves to "get rid" of the others. This may result in projecting the unwanted subself onto outside objects or the dissociation or alienation of parts of the self from the family of selves. This wish to "kick them out of the house" is often mutually felt by various warring subselves. (*c*) Inability to separate one subself from another. This undue fusion of the subselves prevents their expression, development and clear identification. It presents confusion to those more conflict-free subselves whose main function seems to be observation of the total self. (*d*) Superficial, undimensional appreciation of each subself by the others. Each subself fails to see the others "in depth" with integrative, adaptive functions, strength and "goodness." (*e*) All of the above conflicts can be externalized and therefore perpetuated as an interpersonal struggle or a struggle between competing value systems. This struggle can be between the person and his primary objects or between the person and some supposedly alien value system or group pressures. (*f*) The amount of hostility and anxiety generated by the above further interferes and cuts down on internal communication . . . a vicious cycle as in a repressive national state, authoritarian industrial organization or a competitive conflict-ridden family structure. (*g*) Another way of expressing all of the above pathology is in terms of weakness of the synthetic or organizing function of the ego relative to the other parts of the personality.

F. HYPOTHESES ON EGO THERAPY

1. The over-all model of ego therapy is that of a kind of family group therapy within the total self organization.

2. Reconditioning of modal operant responses of the subselves takes place with the patient himself rather than the therapist as the chief source of reinforcement. This is something like the model of a teaching machine in which the patient is both the machine and the subject.

3. One of the aims of therapy is to bring out the fact that there is more ego or self behind most symptoms than most patients (or therapists) realize. The goal here is to bring out the degree and kinds of involvement various subselves have in the symptoms. Far from being simply wild eruptions of the id, the most "irrational" behavior has primary and secondary gains and is often of great adaptive or even survival value for some of the various subselves and the total person. When the patient discovers these primary and secondary gains of the symptoms for the subselves they serve, the patient is likely to feel pride in his symptom or in himself (i.e., operant responses of pride tend to be reinforced at the expense of self-contempt responses). Frequently patients report that they felt helpless before such discovery, but that with this kind of discovery they actually regard the self as strong and in control. "I'm not weak like I always thought. Really it was very clever and strong of me to think up such a system. I even admire myself for it."

4. One of the therapist's most important functions is to facilitate the patient's conscious discrimination learning of the fact that there are different subselves and that these subselves are usually aligned with various objects, including the therapist (e.g., a critical super-ego self aligned with mother, husband and projected onto the therapist).

5. It is also the role of the therapist to help the patient identify some of the subselves and to describe them or characteristics about them. (e.g., two subselves may be fused and thus hard to identify, or there may be an oscillation between two subselves or one subself is completely rejected by the other).

6. The therapist should take a position equidistant from the various subselves, if possible, because reinforcement of one subself at the expense of another prevents the patient from experiencing the *validity* of all the subselves.

7. Isolation in various forms (feelings vs. thoughts, "shoulds" vs. "want to's," subself #1 vs. subself #2 etc.,) is subjectively experienced as being

alienated from oneself or being split into parts—not being a "whole person." This kind of subself isolation is highly correlated with isolation from other people, including the therapist and the experience of therapy itself. The general methodology of the therapist here is to help the patient bring the various subselves together by first separating them enough to be clearly identified. The therapist then ideally provides a model for the acceptance and free interchange of all the subselves.

8. It is especially important to facilitate the patient's ability to experience one or more subselves vis-à-vis the therapy or the therapist in order to reveal the predominant subself alignment which the patient has made with therapy and the therapist. For example, does the therapist or therapy line up with childhood "id-like" subselves or is therapy most often associated with a masochistic self or a critical "super-ego-like" self etc.? It is important for both patient and therapist to realize that therapy is usually experienced on some levels as a deep humiliation by at least one of the subselves.

9. The various strategic interpersonal maneuvers of the type stressed by the Sullivanians should be pointed out to the patient as understandable and persistent attempts to enlist the therapist in the patient's "civil war," on one side or the other, for the purpose of externalizing and perpetuating the conflict. If this kind of externalization occurs, the situation involves any given subself (e.g., subself #1) *plus* the therapist as being against subself #2, and soon the therapist vs. subself #2. This kind of one-sided involvement is something the therapist should be aware of as an ever present temptation to serve his own civil war—as it so often serves the inner struggles of other significant people in the patient's life. That is, the therapist adopts a position equidistant from the various subselves, despite the attempts of the patient to enlist him on one side or the other.

10. Although the interaction between patient and therapist is very significant as indicated above in hypotheses 4, 7, 8 and 9, the most crucial movement is not between therapist and patient. It is between the shifting subselves which move toward, away from or against one another and take on different components as they interact and change. Although it should be made clear which of the subselves in the patient become aligned with the various subselves in the therapist, it is most important that the synthetic function of the therapist's ego provide a model and reinforcing agent for the synthetic function of the patient's ego.

11. One therapeutic aim is to reduce the destructive conflict among the various subselves of the patient. This does not involve total elimination of inner conflict as some inner conflict is regarded as both desirable and in-

evitable—just as between people. Various types of resolution of destructive subself conflicts may result when, for example, a child-like subself is brought into direct and open contact with super-ego-like subself and "learns" that this parental subself "really cares for the child subself." Or this "war" may be resolved in favor of the parental subself in which case the child subself "makes peace" with the super-ego by apology, atonement, understanding, etc. The important point here is that the therapist helps the patient to see various types of peace-making alternatives and their advantages, disadvantages and "costs" to the total personality.

G. SUMMARY

The major purpose of this paper has been to present a theoretical position which combines certain aspects of the structured and unstructured poles in psychotherapy. Based on a modification of both the structural hypothesis of psychoanalysis, and certain propositions from the ego psychology of Hartmann et al. (7), it retains the idea of technique over and above "being oneself" at the same time it allows for a non-transference centered modeling of the therapist's personality for the patient—especially reinforcement of the synthetic function of the patient's ego.

After a discussion of the dilemmas of certain structured and unstructured positions, some hypotheses on ego organization, ego pathology and ego therapy were proposed to resolve some of these dilemmas. The personality is seen as a system of subselves under a more or less well integrated identity. The pathology is mainly conflict between these subselves and the therapy, a kind of family group therapy, within the self organization.

REFERENCES

1. BARKER, E., & SUNDLAND, D. M. The Methods and Attitudes of Psychotherapists—A summary of major results thus far. Report to Respondents, August 1960 (mimeographed reports).
2. FENICHEL, O. Problems of Psychoanalytic Technique. Albany, N.Y.: 1941. Pp. 122.
3. FREUD, S. Papers on Technique. Coll. Papers II. London: Hogarth, 1953. Pp. 402.
4. ———. The Ego and the Id. London: Hogarth, 1927. Pp. 88.
5. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409-416.
6. HARTMANN, H. Technical implications of ego psychology. *Psa. Quart.*, 1951, **20**, 31-43.
7. ———. Ego Psychology and the Problem of Adaptation. New York: International Univ. Press, 1958. Pp. 108.
8. KRIS, E. Ego psychology and interpretation in psychoanalytic therapy. *Psa. Quart.*, 1951, **20**, 15-30.
9. LOEWENSTEIN, R. M. The problem of interpretation. *Psa. Quart.*, 1951, **20**, 1-14.

10. MENNINGER, K. Theory of Psychoanalytic Technique. New York: Basic Books, 1958. Pp. 179.
11. PASCAL, G. R. Behavioral Change in the Clinic—A Systematic Approach. New York: Grune & Stratton, 1959. Pp. 120.
12. ROGERS, C. R. Client-centered Therapy. New York: Houghton Mifflin, 1951. Pp. 560.
13. SHAW, F. J. The reconciliation of contradictions as an integrating concept. ETC., 1961, 18 (in press).
14. WAELDER, R. Basic Theory of Psychoanalysis. New York: International Univ. Press, 1960. Pp. 254.

Western Psychological Center
4911 Van Nuys Boulevard
Sherman Oaks, California

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 91-104.

CONVERGENCE AS A DETERMINER OF PERCEIVED ABSOLUTE SIZE*[1]

U.S. Army Medical Research Laboratory, Fort Knox, Kentucky

WALTER C. GOGEL

A. INTRODUCTION AND PROBLEM

There is evidence that the convergence of the eyes is often inadequate in determining *S*'s perception of the distance of an object from himself (2, 3). Also, perceived size and perceived distance as expressed by the size-distance invariance hypothesis (9) are usually regarded as related. It is therefore somewhat surprising to find considerable evidence for the conclusion that the convergence of the eyes can determine perceived size. This evidence is of several types. The first type of evidence results when false binocular fusion, as a consequence of the incorrect convergence of the eyes, occurs with respect to a repeating pattern of visual stimuli. It is found, for example, that the overconvergence to a recurring design (a design such as sometimes occurs on wallpaper) will make the design appear smaller than normal. A number of instances illustrating this same basic phenomenon have been reported (10, pp. 425-432; 13, pp. 676). A second type of evidence relating convergence and perceived size involves changing the convergence to a binocular stimulus presented in a stereoscope or as after-images. It is found that, under these circumstances, the perceived size of the object varies inversely with the convergence (6, pp. 312-313; 11; 13, pp. 674-676). A third type of evidence results from presenting two objects at two different convergence values and having *S* adjust (or select) the size of the one object until it appears to be the same as the other. This is usually repeated with different amounts of convergence difference between the two objects. In this type of study it is essential that all extraneous cues be eliminated, and that the two objects be presented successively not simultaneously. If this latter criterion is not met, binocular disparity will be present and the results cannot be attributed to the factor of convergence. From studies meeting these requirements (5, 7, 8) it can be concluded that the perceived ratio of the sizes of objects changes as a function of their convergence difference.

---

* Received in the Editorial Office on September 6, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to thank Charles A. Wainwright, Jr. for his assistance in collecting and analyzing the data of this study.

It is clear, therefore, that convergence and perceived size are related. From the first type of study, it might be concluded that convergence can determine how large an object appears to be. From the second type of study, it might be concluded that changes in convergence can determine changes in the perceived size of this object. From the third type of study, it might be concluded that convergence differences can determine the perceived relative sizes of objects.

These conclusions are concerned with two kinds of convergence. One kind is absolute convergence while the other kind consists of convergence differences. Two kinds of perceived size are also involved. These are perceived absolute size and perceived relative size. The absolute convergence of the eyes is the angle between the axes of the eyes, when the eyes are so positioned that the image of the object is on corresponding points on the two foveas. A convergence difference is measured by the magnitude of the change between two such positions of the eyes when looking from one distance to another. A perception of relative size is the perceived ratio of the apparent sizes of the two objects. Perceived absolute size concerns the perceived size of the object regardless of its perceived size with respect to other objects.

One possible explanation of the relation between convergence and perceived size is by means of the concepts of absolute convergence and perceived absolute size. Suppose that an absolute convergence determines a perceived absolute size as the first type of evidence seems to suggest. Then the process by which a convergence change or difference determines a change in perceived size or a ratio of perceived sizes is immediately clear. In this case the relational type of size judgment is merely a consequence of two absolute size judgments resulting from the several absolute values of convergence. If on the other hand, it is found that absolute convergence does not determine perceived absolute size, some other explanation for the relational type of judgment must be invoked. Thus, it is important to know whether absolute convergence can determine perceived absolute size. The present study will be concerned with a direct test of this possible relation.

In the present study, perceived absolute size was measured by having *S*s make a kinesthetic (non-visual) judgment of the size of a single, binocularly presented object. This object had a constant angular size and was presented at each of several convergence distances. It is clear that the physical size of an object producing a constant angular size would increase linearly with distance. Therefore, if convergence is a cue to perceived absolute size, the perceived size of the object should be greater with the lesser convergence values.

Since only one visual object was present with the successive presentations being separated by reasonably long time intervals, the size judgments presumably did not involve a comparison of several visual objects. In this sense, only absolute size judgments occurred. For the same reason, convergence per se (absolute convergence), rather than a convergence change, was presumably the only possible size cue available.

B. METHOD

1. *Apparatus*

a. The binocularly observed object. Five discrete values of convergence (24, 78, 132, 186, and 240 minutes of arc) were used. The object whose size was to be judged at each of these convergence values was a stereoscopically generated object of a constant angular size of 2°54′20″. The *S*'s view of this object in each eye and his view of the object when it was binocularly fused is illustrated in Figure 1. This stereoscopic object is called the stake reticle. It is seen from Figure 1 that the triangular pips on the stake reticle were viewed monocularly while the remainder of the stake reticle was viewed binocularly. The purpose of the monocular pips was to assure that the binocular presentation was properly fused. When the binocular view of the reticle was that shown in the "fused" view of Figure 1, the convergence of the *S*'s eyes was to the required value.

The binocular stake reticle was generated by a pair of reflex sights, with one reflex sight before each eye of the *S*. Each reflex sight contained an aperture source of light of the particular shape required for the stake reticle. The light from this source passed through an optical system and emerged as parallel light. This parallel light was then reflected from a reflecting-transmitting mirror into an eye of *S* as the luminous stake reticle. The principles involved in this instrument are discussed in a previous study involving the generation of a stake reticle (1).

By turning an adjustment knob, *E* could rotate the reflex sights simultaneously in opposite directions. This changed the lateral position of the stake reticle in opposite directions in each eye. Thus, by adjusting the instrument knob to the proper position, *E* could specify the convergence position of the eyes which would occur when *S* fused the binocular object. A switch on the instrument permitted the reticle light sources to be turned off while *E* was adjusting the convergence of the reticles.

The reflecting-transmitting mirrors which directed the reticles into *S*'s eyes were such that real objects placed in front of *S* as well as the binocular reticle could be observed with both eyes when a black velveteen curtain at the front of the instrument was raised. A black hood could be adjusted over *S* so that when the black curtain was lowered and the hood was adjusted, nothing was visible to *S* except the binocular stake reticle.

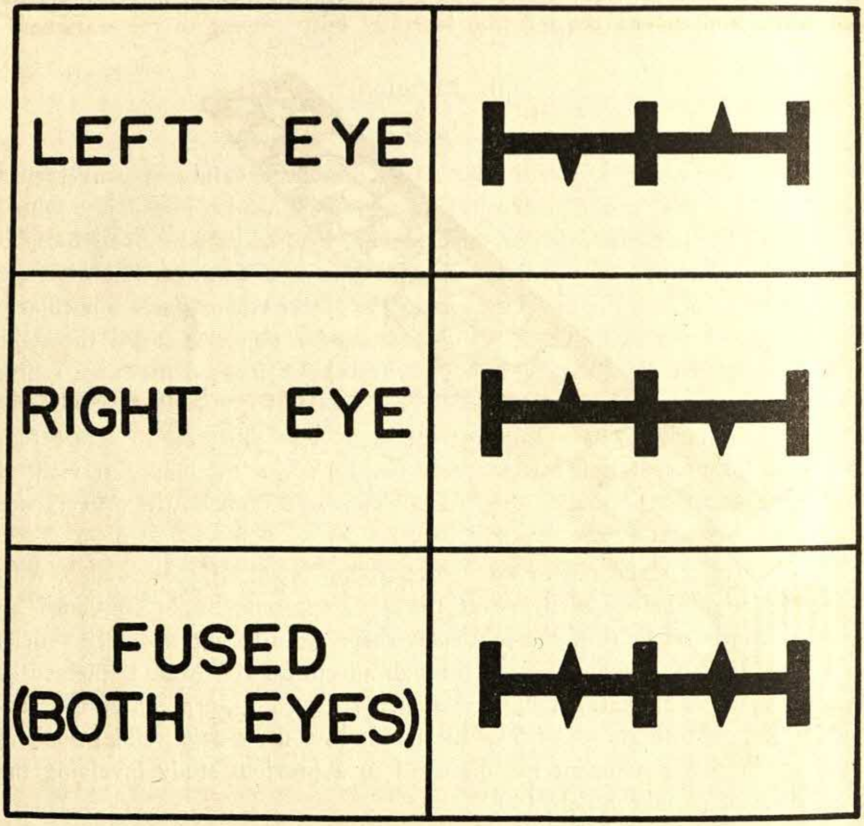


FIGURE 1
A SCHEMATIC REPRESENTATION OF THE RETICLES USED TO GENERATE THE BINOCULAR OBJECT

The accommodation to the binocular stake reticle was always the same as that required for a very distant object regardless of the convergence value. This had the advantage of holding accommodation constant while testing the relation between convergence and perceived size. It had the disadvantage of limiting the amount of convergence which could be used in the experiment. At large values of convergence the stake reticle would tend to become in-distinct due to the dissociation of the normal accommodative-convergence relationship.

b. The kinesthetic adjustment. A pair of aluminum rods were vertically positioned at about the level of *S*'s waist. The *S* grasped these rods, one in each hand, and moved the left rod laterally with respect to the stationary right rod until the lateral distance between the centers of the rods was judged to be the same as the width of the binocular stake reticle. The rods were invisible to *S* during the adjustment. A measuring stick attached to the movable rod enabled *E* to measure the magnitude of the kinesthetic width adjustments.

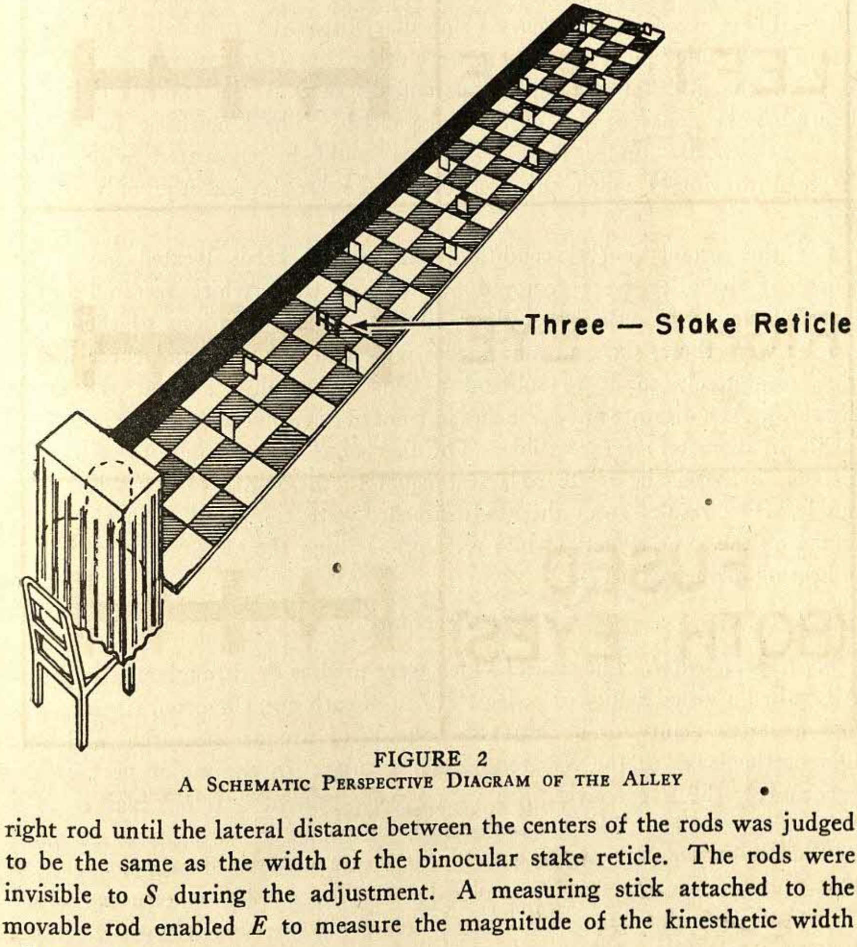


FIGURE 2
A SCHEMATIC PERSPECTIVE DIAGRAM OF THE ALLEY

c. Controls. Two control conditions were used. The purpose of these controls was to demonstrate that the kinesthetic adjustment method constituted a sensitive and reliable method of measuring perceived size. For one control, the curtain in front of the stereoscopic instrument was raised revealing an alley extending away from *S*. A schematic perspective view of this alley is shown in Figure 2. The alley was 34.2 feet long and 4.2 feet wide and contained a number of vertical white cards distributed uniformly in distance along the alley. The entire alley as well as the stake reticle was viewed binocularly. The stake reticle appeared to be suspended above the alley. There were effective cues (binocular disparity) to localize the stake reticle in the alley in depth. There were other cues (the apparent size of the parts of the alley at the same apparent distance as the stake reticle) to determine the apparent size of the stake reticle. The kinesthetic judgments of perceived size under these conditions could be compared with those obtained previously when the binocular stake reticle was the only object visible.

For the second control condition, the vertical cards located along the center of the alley were removed and one of three white rectangles was placed facing *S* at either 10, 20 or 30 feet. The dimensions of these rectangles were three, six, and nine inches wide and one, two and three inches high respectively so as to subtend a constant angular size at the three distances. At a distance of 46.5 inches in front of *S*, two vertical white rods (0.2 inches in diameter) were visible. The task of *S* was to indicate when the distance between the white rods was equal to the width of the white rectangle. The results from these adjustments were compared with those obtained on the same targets (white rectangles) using the kinesthetic method of indicating perceived size.

2. *Procedure*

a. Experimental. The same 12 men were used as *S*s throughout the study. All *S*s had a visual acuity of at least 20/30 in each eye, corrected if necessary, and a stereo acuity of at least 85 seconds of arc as measured with the Diagnostic Series of the Keystone Telebinocular. In the major part of the experiment, the five convergence values were presented four times to each *S* in random order. The *S* was carefully instructed as to his task. A small (4 inch wide) model of the stake reticle was used to illustrate the final binocular configuration which was required (the fused condition in Figure 1). Following this, the *S* was seated before the instrument with the front curtain closed and with the black hood adjusted in position. Under these conditions, *S* was in complete darkness. The *E* adjusted the instrument dial to produce the required value of convergence, the reticle sources were turned on, *S* fused the binocular reticle and kinesthetically adjusted the distance between the aluminum rods to apparently reproduce the right-left extent of the stake reticle. Following this, the reticle sources were turned off, the instrument knob was adjusted to another convergence value and the process was repeated until all 20 presentations were completed. The time interval between presentations of the different convergence values was approximately one minute. This portion of the experiment is called the "reduced cue situation" since only the binocular stake (not the alley) was visible.

b. Controls. After completing the adjustments under the reduced cue conditions *S* was presented with the first control situation. The procedure here was identical to that used in the first portion of the experiment except that following the binocular fusion of the stake reticle the curtain in front of the stereoscopic instrument was raised revealing the lighted alley. The *S*, as before, kinesthetically adjusted the distance between the aluminum rods to the same apparent extent as the width of the stake reticle. The situation in which the binocularly observed alley as well as the binocular reticle was visible is called the "full cue situation."

In the second control condition, both kinesthetic and visual adjustments were made to the apparent size of the targets at 10, 20 or 30 feet. The kinesthetic adjustments as before consisted of adjusting the distance between the aluminum rods by hand to apparently reproduce the target width. The visual adjustments consisted of *S* signalling when the apparent distance between the visible rods in the alley seemed to reproduce the target width. In this control the binocular reticle was not present. Throughout the experiment the instructions were such as to encourage objective matching i.e., the task of *S* was to reproduce the physical width of the target object.

C. RESULTS

1. *Experimental*

The average results from the portion of the experiment using the binocular reticle are given in Figure 3. The curve in Figure 3 labelled "perfect constancy" gives the physical widths at the several convergence distances which would be required in order to produce the constant 2°54′-20″ angular size. For example, if perfect size constancy had been present, the binocular reticle should on the average have been perceived as being about 48 cm. wide at the 24′ convergence and about 5 cm. wide at the 240′ convergence.

The curve labelled "reduced cue situation" gives the average results from the 12 *S*s for the situation in which the binocular stake but not the alley was visible. Under this condition only convergence was a possible determiner of perceived size and it will be seen that only a very small average change in perceived size occurred as a function of convergence. A comparison of the reduced cue situation curve and perfect constancy curve of Figure 3 indicates how little size constancy occurred when only convergence was available as a determining cue.

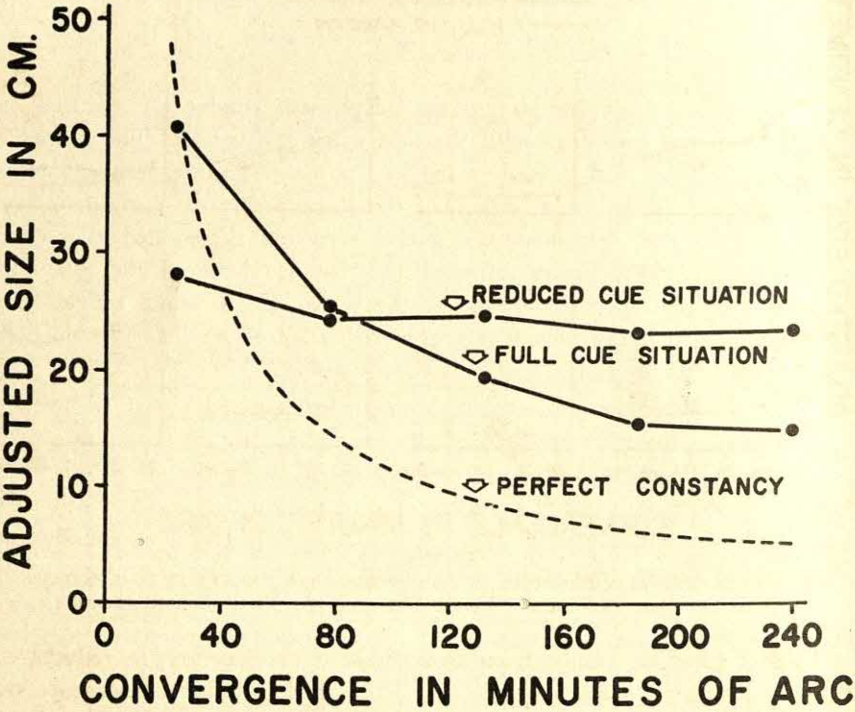


FIGURE 3
PERCEIVED AND PHYSICAL SIZE OF AN OBJECT OF CONSTANT ANGULAR SIZE AS A FUNCTION OF CONVERGENCE

The solid line curves of Figure 4 give the average results from each *S* for the "reduced cue situations." The letters in Figure 4 represent a method of identifying the *S*s common to this and previous studies (2, 3). An examination of the solid line curves of Figure 4 reveals that approximately half the *S*s showed some evidence of a change in perceived size as a function of convergence. But, even for these *S*s, the change in perceived size was small.

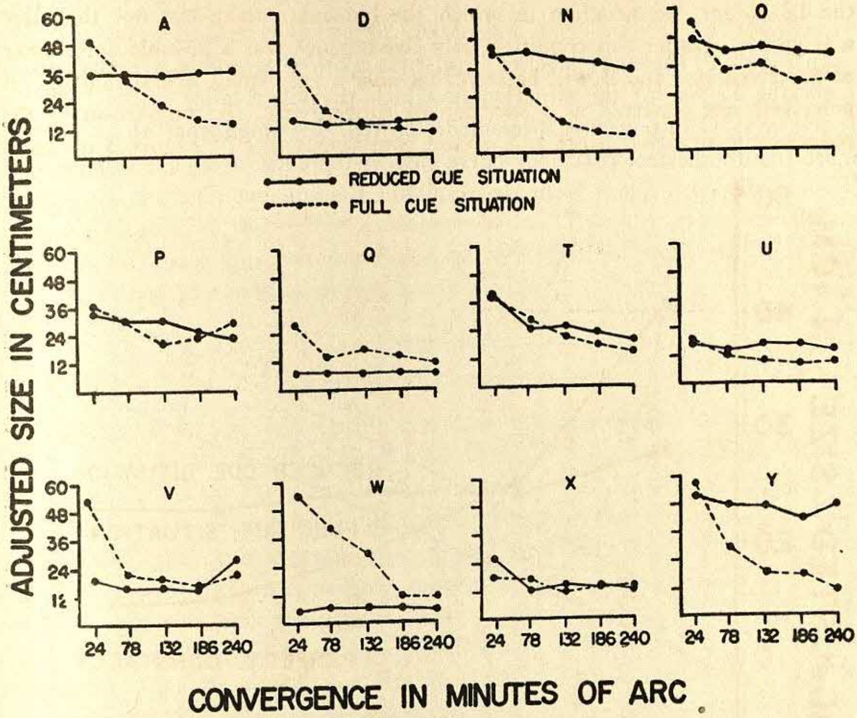


FIGURE 4

Perceived Size as a Function of Convergence for an Object of Constant Angular Size for Each *S* Under Reduced and Full Cue Conditions

Table 1 gives the results from an analysis of variance for the reduced cue situation using the data from all *S*s. Table 1 shows that convergence was a statistically significant variable in determining perceived size. Considering Table 1 and Figures 3 and 4, it is clear that a significant but small amount of change in perceived size sometimes occurred as a function of convergence in the reduced cue situation.

TABLE 1

Summary of Analysis of Variance of "Reduced Cue" Data

Source of variation	*df*	Mean square	*F*
Convergence	4	46.30	4.15*
Subjects	11	978.14	87.73*
Error	44	11.15	
Total	59		

* Significant at the .01 level.

2. *Controls*

In Figure 3, the curve labelled "full cue situation" gives the average results from the 12 *S*s for the situation in which the visual alley as well as the reticle was simultaneously presented. It will be noted that this curve is more like the perfect constancy curve than is the curve from the reduced cue situation. As is evident from the results of the full cue situation, the kinesthetic method of adjustment can discriminate between different perceived sizes. The differences in shape between the full cue curve and the perfect constancy curve is attributed to the three-stake reticle having been located in direction considerably above the floor of the alley and thus somewhat removed from the objects located along the floor of the alley. The average results from each *S* for the "full cue situation" are given by the dashed line curves of Figure 4.

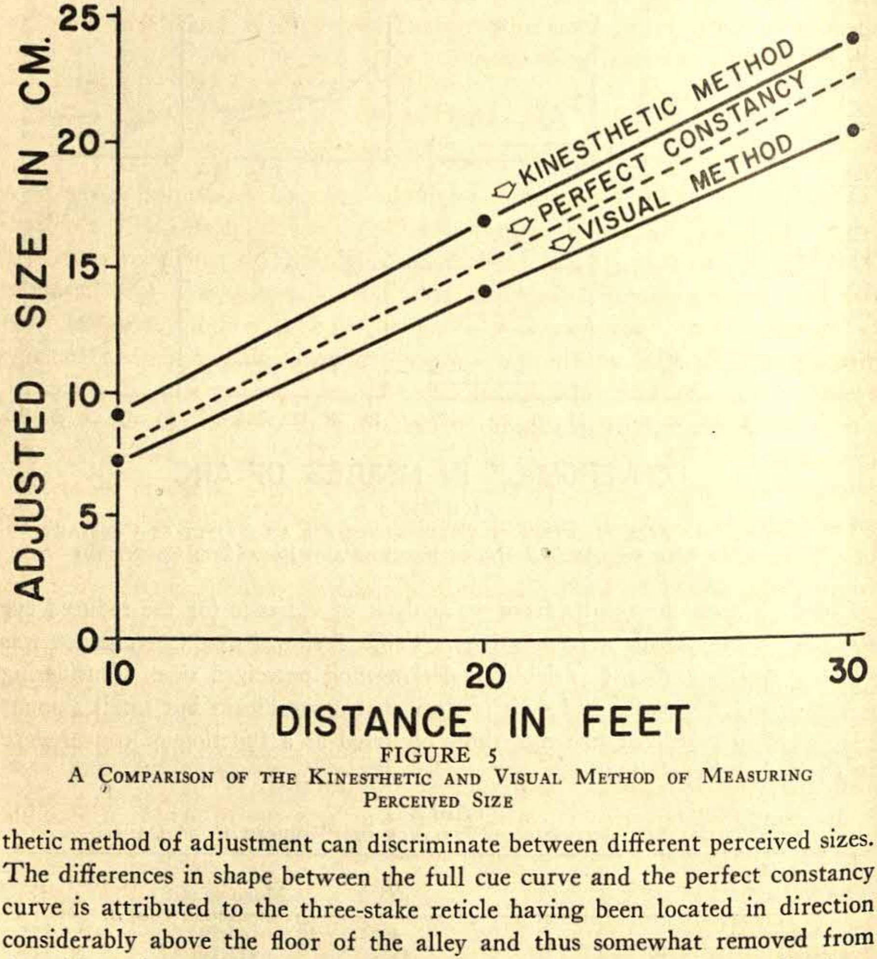


FIGURE 5

A Comparison of the Kinesthetic and Visual Method of Measuring Perceived Size

The average results from the 12 *S*s for the kinesthetic and visual judgments of the width of the white rectangles located in the alley at 10, 20 and 30 feet are given in Figure 5. The curve labelled "perfect constancy" gives the results which would have been obtained if the physical widths of the rectangles at the different distances had been judged correctly. It will be noted that the slopes of the kinesthetic and visual curves are similar and are close to that of the perfect constancy curve. Evidently, the kinesthetic adjustment method can measure perceived size with as much sensitivity as can the method of visual adjustment.

D. Discussion

1. *Theoretical Considerations*

One way in which convergence might be expected to determine the perceived size of an object is by determining the perceived distance of the object. Thus, an object of a particular retinal size might have a particular perceived size as a consequence of being seen at a particular distance. Or, it might be supposed that convergence and retinal size determine perceived size directly without the intermediary process of perceived distance. In any event, retinal size and convergence must interact perceptually if perceived size is to be determined in a reduced cue situation such as that used in the present study. The reason for this is clear. If absolute convergence determined perceived size independently of retinal size, this would mean that an object at a particular convergence distance would be perceived as being of a particular size regardless of its angular or physical size. In this case, convergence would be an irrelevant and nonadaptive cue system since the resulting perceived sizes would have no relation to the actual physical sizes of the objects. Clearly, convergence, if it is to be adaptive, cannot by itself be a determiner of perceived absolute size. If, however, retinal size can be assumed to contribute to the perception of absolute size, then absolute convergence and retinal size together could determine perceived size in a manner such that the perceptions could approach veridicality.

It seems therefore that if convergence is to be a cue to perceived absolute size in reduced cue situations, then retinal size must have perceptual consequences. The retinal size referred to here is an absolute retinal size, e.g., a retinal size in millimeters. It is absolute in the sense that it involves a retinal extent and not a change in retinal extent. It is asserted that there is evidence for rejecting the hypothesis that absolute retinal size has perceptual consequences and hence for rejecting the hypothesis that absolute convergence is a determiner of perceived absolute size.

One portion of the evidence against the absolute size of the retinal image having perceptual consequences is given in a study by Gogel, Hartman and Harker (4). In this study, it was found that the angular size of a familiar object was not a determiner of perceived absolute distance. This result would be expected only if the angular (retinal) size of the familiar object had no perceptual consequences. Another study relevant to this problem is a study by Wallach and McKenna (12) in which it was demonstrated that the perceived absolute size of a retinal stimulus in a reduced cue situation bore no fixed relationship to the perceived size of objects under more normal environments. Thus, it is unlikely that the absolute retinal size of a visual stimulus has perceptual consequences. It follows that it is also unlikely that absolute convergence is a cue to perceived absolute size.

2. *Interpretation of Results*

It is clear that the method of kinesthetic adjustment is a sensitive indicator of perceived size. Thus, the present study suggests that between 24 and 240 minutes of arc, absolute convergence has little effect upon perceived absolute size. In linear distance, for the *S*s of this experiment, this is an average range from 941 to 94 centimeters. It appears from previous studies (5, 8) that convergence values within this depth interval can determine changes in perceived relative size. However, the results of the present study question whether such changes in perceived relative size can be considered to be the consequence of changes in perceived absolute size, as a function of different values of absolute convergence.

But, even though convergence cannot be considered as having been very effective in the present study, as is indicated by Table 1, some amount of change in perceived absolute size as a function of convergence did occur. It is possible, however, that these changes in the perceived size can be explained by relative size judgments occurring between the successive visual presentations. This would mean that the judgments of size which were made with respect to the first presentations would be largely haphazard. The response to the first stimulus would be more or less random. The response to the second and all succeeding stimuli however, would be made relative to those occurring earlier. If this explanation is correct, the small amounts of apparent size change, which sometimes occurred with the different convergence values in the reduced cue situation, would be a relative change in perceived size resulting from a convergence difference, with the time interval between different convergences being rather large.

If the first response in the reduced cue situation was random with the following responses based upon it (as is required by the above argument) the *S*s would be expected to differ in their average size judgments. Some evidence for this is given in Table 1 which shows that most of the variance in the experiment can be attributed to differences between *S*s independent of the changing convergence. It seems that an *S* in the reduced cue situation tended to respond with a size estimate which was different from that of other *S*s. Also, the responses from different *S*s tended to differ consistently throughout the reduced cue situation. This is in agreement with the contention that the first response was randomly determined with each succeeding response being made in relation to preceding ones. Thus, it is doubtful whether the small changes in perceived size as a function of convergence, which occurred in the reduced cue situation, can be considered to be changes in perceived absolute size. More likely they should be characterized as changes in perceived relative size with the relative judgments being made between successive presentations.

E. Summary and Conclusions

The possible relation between convergence and perceived absolute size was determined for convergence values of 4° or less with accommodation held constant. Perceived absolute rather than perceived relative size was measured by having *S* reproduce kinesthetically the perceived size of the visual stimulus. The visual stimulus consisted of a stereoscopic object of constant angular size, the convergence to which was changed by *E* in discrete steps. The results suggest that about half the *S*s perceived some change in size with a change in convergence. But, even with these *S*s, the change was small and possibly can be attributed to comparisons between successive presentations. In general, the results of this study indicate the inadequacy of convergence as an effective cue to perceived absolute size.

The sensitivity of the kinesthetic method of indicating perceived size was tested by comparing it with a visual method. The sensitivity of the two methods of measuring size constancy was found to be similar. The validity of the kinesthetic method was indicated in that the addition of distance cues to a visually reduced situation resulted in a change in the kinesthetically measured size in the direction of object constancy.

References

1. Gogel, W. C. An observer constant in the perception of stereoscopic depth. USAMRL Report No. 316, 1958.
2. ———. Convergence as a cue to absolute distance. *J. of Psychol.*, 1961, **52**, 287-301.

3. ———. Convergence as a cue to the perceived distance of objects in a binocular configuration. *J. of Psychol.,* 1961, **52**, 303-315.

4. GOGEL, W. C., HARTMAN, B. O., & HARKER, G. S. The retinal size of a familiar object as a determiner of apparent distance. *Psychol. Monog.,* 1957, **70**, 1-16.

5. HEINEMANN, E. G., TULVING, E., & NACHMIAS, J. The effect of oculomotor adjustments on apparent size. *Amer. J. Psychol.,* 1959, **72**, 32-45.

6. HELMHOLTZ, H. Treatise on Physiological Optics. Vol. III (J. P. C. Southall, *Ed.*) Menasha, Wisconsin: Banta, 1925.

7. HERMANS, T. G. Visual size constancy as a function of convergence. *J. Exper. Psychol.,* 1937, **21**, 145-161.

8. ———. The relationship of convergence and elevation changes to judgments of size. *J. Exper. Psychol.,* 1954, **48**, 204-208.

9. KILPATRICK, F. P., & ITTELSON, W. H. The size-distance invariance hypothesis. *Psychol. Rev.,* 1953, **60**, 223-231.

10. LINKSZ, A. Physiology of the Eye. Vol. 2, Vision, New York: Grune & Stratton, 1952.

11. TAYLOR, F. V. Change in size of the after-image induced in total darkness. *J. Exper. Psychol.,* 1941, **29**, 75-80.

12. WALLACH, H., & MCKENNA, V. V. On size perception in the absence of cues for distance. *Amer. J. Psychol.,* 1960, **73**, 458-460.

13. WOODWORTH, R. S. Experimental Psychology. New York: Holt, 1938.

Civil Aeromedical Research Institute
Federal Aviation Agency
Oklahoma City, Oklahoma

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 105-118.

AFTEREFFECT OF ELECTRICAL STIMULATION TO THE EYE*[1]

The Training School at Vineland, New Jersey

JOHS. CLAUSEN AND RATHE KARRER

A. PROBLEM

Whether or not repeated determination of electrical phosphene thresholds results in gradually increasing values seems to be a matter of some controversy. Riggs, Cornsweet and Lewis (1957) conclude that no long-lasting changes have appeared in their experiments. Kohata, Komatsu and Motokawa (1956), on the other hand, report that repeated high-intensity electrical stimuli produce long-lasting elevations in the electrical threshold. In a previous publication by the senior author (Clausen, 1955, p. 40) we cited some preliminary results from two observers when ten successive phosphene threshold were determined. The thresholds of one observer increased from the first to the fifth determination with a subsequent levelling off. The other observer showed no sign of increase throughout the series. In a study concerned with age and sex differences (Clausen and Karrer, 1959) we determined five successive thresholds for 142 observers and were impressed by the apparent systematic increases which resulted from this successive determination. We regarded this problem of sufficient importance to justify some closer investigation. The problem has a practical aspect in that to insure accuracy of a threshold determination one might record several successive thresholds. It is also a problem of considerable theoretical interest in that it may clarify the action of an electrical stimulus to the eye.

B. PROCEDURE

We have tried to explore the problem of aftereffect of electrical stimulation in the following ways: (*a*) analyzing the data from the previously cited age-sex study for aftereffects; (*b*) varying the time interval between each stimulus (Experiment I); (*c*) varying the time interval between each successive threshold determination (Experiment II); (*d*) comparing the threshold before a three-minute stimulation to a threshold obtained at varying

---

* Received in the Editorial Office on September 11, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was supported by National Science Foundation Grants G-2106 and G-7029.

intervals after this stimulation (Experiment III); (*e*) determining how many times a stimulus just above threshold had to be presented before it ceased to elicit the visual sensation of flickering light. This method was abandoned because of the difficulties the observer had in giving answers with any degree of confidence and consistency.

As a source of electrical stimulation we employed a Hewlett-Packard Low Frequency Oscillator (Model 202 C). The sine wave signal was applied through two silver electrodes 25 mm. in diameter attached to the middle of the forehead and the right temple. The skin of these areas was cleaned with acetone and the electrodes covered with bentonite electrode paste and fastened with adhesive tape. An interval timer allowed a stimulus duration of exactly one second. The threshold was recorded in μA by means of a Hewlett-Packard Electronic Voltmeter (Model 400 B) with a 1000 ohm shunt resistor. Inserted in the circuit was a 10,000 ohm resistor in series with the observer in order to reduce the effect of fluctuation in the observer's skin resistance.

The thresholds were obtained by using the psychophysical method of limits. The average of an ascending and descending series constituted a threshold. We did not make any attempt to achieve equal steps between stimuli in these series. Rather the steps were larger in the beginning of the series, and smaller as we approached the threshold. Our procedure here is more in accordance with Motokawa than with Riggs.

In our standard procedure the interval between successive stimuli is determined by the time it takes for the observer to report his judgment, plus the time it takes to adjust the dial for the following intensity, plus ca. two seconds between the ready signal and the next stimulus. In total this amounts to about four to five seconds between stimuli. In addition we allowed about five seconds between successive determinations of the thresholds. In the experiments where interval between successive stimuli, or interval between successive threshold determination was varied, we used a modification of this standard technique for stimulus-presentation. All thresholds were determined while the observers were facing a 7½ × 11 foot screen, illuminated according to the experimental conditions.

In Experiment I we utilized an additional interval timer which controlled the interval between successive stimuli. Intervals of five and 10 seconds were used. The same interval was used between successive thresholds as between successive stimuli. In Experiment II we used standard procedure with regard to interval between successive stimuli but we allowed intervals of three, five and 10 minutes to elapse between successive thresholds. Ten successive thresholds were determined for each condition.

Previous studies have shown that the phosphene threshold is determined by current frequency and by the state of light adaptation of the eye (Clausen, 1955, 57 ff.). In Experiment I we used two frequencies; 20 and 50 cps and two light conditions; .03 and 30 mL. At each frequency both light conditions were used. Thus we recorded thresholds for four different conditions.

In Experiment II we also used the frequencies of 20 and 50 cps, but only one light condition, 30 mL.

We selected these frequencies for the two experiments because 20 cps gives the lowest threshold, while 50 cps gives a relatively high threshold which may still be determined without appreciable discomfort. It seemed reasonable to assume that the aftereffect would be more pronounced at higher thresholds. The two different light conditions were included on the supposition that the state of light adaptation might influence the aftereffect to electrical stimulation.

Five observers were used in Experiment I, with three repetitions for each condition. The points in Figure 1 are therefore based on 15 determinations. In Experiment II the number of observers and repetitions varied for the conditions and will be indicated in the Result Section.

In Experiment III we adhered to the standard procedure of phosphene threshold determination. We used a 20 cps sine wave and a 30 mL luminosity. For each session we determined a reference threshold which was followed by stimulation with 100 μA for 3 minutes. At 0, 1, 2, 3, 5, 7 and 10 minutes from the end of the prolonged stimulation we started the determination of a second threshold. Our primary aim here was to determine how long it would take for the system to recover its original level of excitability. Fifteen observers took part in this experiment, each having five determinations for each interval. Only five observers, however, were used in the seven- to 10-minute conditions.

On these five observers we ran a control series where the prolonged electrical stimulation was replaced by continued observation of the illuminated screen. Comparison thresholds were obtained at time intervals corresponding to the 0, 1, 2, 3, 5 and 10 minute conditions of the original experiment; each condition being repeated three times. The purpose of this series was to determine the effect of the background photic stimulation on the phosphene threshold.

C. RESULTS

As mentioned above in the age-sex study we had obtained five consecutive thresholds on each of four conditions which constituted a session. Two sessions were run on different days with high reliability. For the present analysis it was decided to use only the data from the second session for the 20 cps—.03 mL condition, which always occurred first in the session and was therefore uncontaminated by previous threshold determinations. In this study a total of 142 observers were included, and our standard procedure for threshold determination was used. The great majority of these observers were untrained. The group means for each of the five threshold determinations are listed in Table 1. An analysis of variance based on the group means is shown in Table 2. Since none of the interactions involving repetition are significant we have pooled them, to obtain an error term for the main effect of repetition. On this basis the main effect of repetition is highly significant. Thus it appears that the threshold is significantly increased after five consecutive threshold determinations. It is interesting to note, however, that the group which has the lowest threshold,

TABLE 1
GROUP MEANS FOR FIVE CONSECUTIVE THRESHOLD DETERMINATIONS FROM THE AGE-SEX STUDY

	N	1	2	3	4	5
Females below age 15	13	14.5	13.9	14.4	14.7	14.5
Males below age 15	14	19.2	20.0	20.8	21.2	22.0
Females above age 15	59	17.7	18.4	19.5	20.2	20.3
Males above age 15	52	20.2	21.9	23.6	25.8	25.9

TABLE 2
ANALYSIS OF VARIANCE OF DATA FROM THE AGE-SEX STUDY

	DF	Sum of squares	Estimate of variance	F
Source				
1. Repetition	4	23.12	5.78	23.78*
2. Sex	1	4.91	4.91	—
3. Age	1	73.35	73.35	—
Interactions				
1 × 2	4	0.42	.11	—
1 × 3	4	5.88	1.47	1.11
2 × 3	1	137.80	137.80	103.61*
1 × 2 × 3	4	5.31	1.33	
Total	19	250.79		

* Significant at .01 level of confidence.

females below age 15, does not show an increase with repetitions. The significant interaction between age and sex is of no concern to the present study.

The results of Experiment I in the present series are shown in Figure 1 which presents the group means for the five observers on whom a series of thresholds were determined by presenting the stimuli five or 10 seconds apart for two light frequencies and two light conditions. Whereas 10 repetitions were made, we selected every other threshold (second, fourth, sixth, eighth and tenth) for an analysis of variance. The analysis of variance is presented in Table 3. In this analysis we are primarily interested in the variance contributed by repetition and time. For all third order interactions we have used the fourth order interaction as error term. Depending upon significance at this level, second order interactions have been checked against the appropriate third order interaction or against pooled third and fourth order interactions. Proceeding in this fashion, we found six

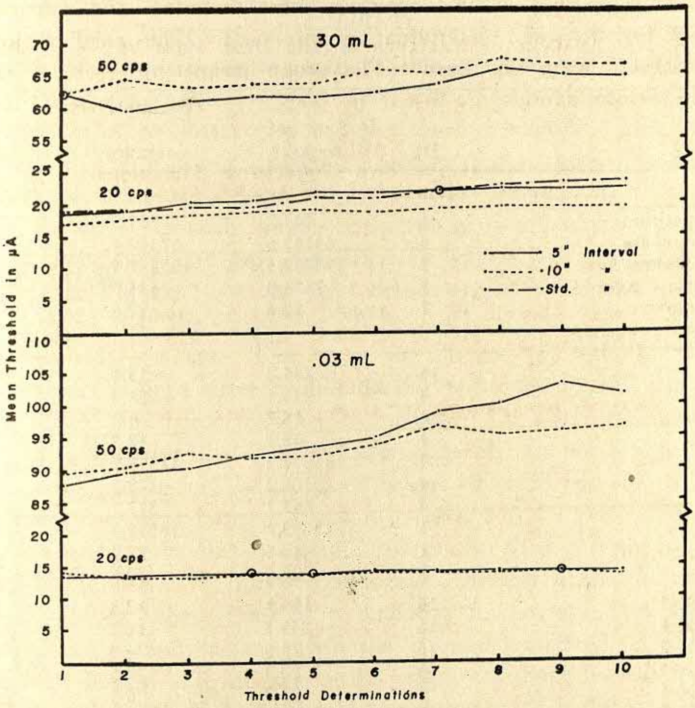


FIGURE 1

Mean Thresholds for Ten Consecutive Determinations significant F-values. Of the main sources, only frequency is significant. This reflects the well established finding that thresholds at 50 cps are higher than thresholds at 20 cps. The interaction Repetition × Frequency is significant, indicating that throughout the series there is a tendency for increased threshold at one frequency, and not for the other but it cannot be determined from the data which one it is.

TABLE 3
ANALYSIS OF VARIANCE OF THE DATA FROM THE INTERVAL BETWEEN THRESHOLD EXPERIMENT

	DF	Sum of squares	Estimate of variance	F
Source				
1. Repetition	4	447.7	111.9	2.08
2. Observers	4	69449.6	17362.4	5.57
3. Frequency	1	195237.5	195237.5	362.2***
4. Light	1	7535.0	7535.0	2.42
5. Time	1	12.6	12.6	—
Interactions				
1 × 2	16	246.2	15.4	—
1 × 3	4	215.8	53.9	4.78**
1 × 4	4	16.7	4.2	—
1 × 5	4	48.8	12.2	1.08
2 × 3	4	46078.9	11519.7	3.70
2 × 4	4	1508.2	379.6	1.22
2 × 5	4	868.3	217.1	—
3 × 4	1	16839.8	16839.8	5.41
3 × 5	1	32.7	32.7	—
4 × 5	1	20.6	20.6	—
1 × 2 × 3	16	199.9	12.5	1.46
1 × 2 × 4	16	259.5	16.2	1.89*
1 × 2 × 5	16	78.6	4.9	1.39
1 × 3 × 4	4	82.3	20.6	2.41
1 × 3 × 5	4	25.2	6.3	—
1 × 4 × 5	4	4.5	1.1	—
2 × 3 × 4	4	1246.0	311.5	8.56*
2 × 3 × 5	4	967.7	241.9	6.65*
2 × 4 × 5	4	200.1	50.0	1.37
3 × 4 × 5	1	146.3	146.3	4.02
1 × 2 × 3 × 4	16	171.7	10.7	1.34
1 × 2 × 3 × 5	16	104.2	6.5	—
1 × 2 × 4 × 5	16	158.0	9.9	1.23
1 × 3 × 4 × 5	4	19.5	4.9	—
2 × 3 × 4 × 5	4	745.6	36.4	4.55*
1 × 2 × 3 × 4 × 5	16	128.6	8.0	

* Significant at $<$ .05 level of significance.
** Significant at $<$ .01 level of significance.
*** Significant at $<$ .001 level of significance.

The significant interaction between Repetition × Observers × Light shows that repetition contributes significant variance for some observers at either light condition. Inspection of the data indicates that the threshold for some observers is increased by repetitions at the .03 mL light condition, while the threshold for other observers in increased at 30 mL.

From the significant interaction of Observers × Frequency × Time it appears that time contributes significantly to variance in some observers at either frequency. Inspection of the data shows that there is a general trend for the threshold at 20 cps to be higher for the five-second interval, while under the 50 cps conditions some observers' thresholds are higher at 10 seconds while other observers are higher at five seconds. The interaction Observers × Frequency × Light does not interest us here. No further information can be obtained from the higher order interactions.

Since we were primarily interested in the main effects of repetition and time and since these are only significant in interaction with other factors, the curves in Figure 1 are the most descriptive of tendencies in the data. It can be seen that the curves for 20 cps—.03 mL for both five- and 10-second intervals are flat. The curve for 20 cps—30 mL for 10-second intervals shows a slight increase while the curves for five seconds and standard interval show a consistent tendency to increase. The curve for the standard interval is taken from the data in Experiment II. Its correspondence with the five-second curve will be discussed below. The 50 cps—.03 mL curves both show a consistent increase which is greatest for the five-second interval. It can be seen that the 50 cps—30 mL curves also show a slight increase but considerably less than the increase at .03mL. Finally it may be noted that in the conditions where there is a tendency for increase, the five-second interval shows greater increase than the 10-second interval. The only exception is at 50 cps—30 mL where the initial threshold for the five-second interval is fortuitously high. It should be pointed out that the most consistent increase for 20 cps is at 30 mL, while for 50 cps it is at .03 mL.

The essence of this is that even in trained observers there may be stimulus conditions which result in a gradual decrease of sensitivity unless some efforts are made to prevent it. One such measure is spacing the stimuli. Another possibility is to pause between threshold determination to allow restitution of the system.

Experiment II in our study deals with the interval between thresholds. The primary purpose was to determine whether an interval between thresholds was sufficient to dissipate any aftereffects resulting from our standard procedure. The data of this experiment did not lend themselves to analysis of variance. Instead, we compared the first and last threshold of a series in terms of C. R. These comparisons are presented in Table 4. From this table it can be seen that a three-minute interval between thresholds still gives a significant rise throughout the series. With five or 10 minutes between thresholds no significant rise occurs. Thus an interval of three to five minutes between thresholds is necessary to prevent a build up. Obviously the spacing of stimuli is more efficient. A rapid repetition of stimuli seems to decrease the excitability more profoundly, with a correspondingly longer recovery time. It appears that we have marked temporal summation effects of the electrical stimulus.

TABLE 4
COMPARISON OF FIRST AND LAST THRESHOLD OF
INTERVAL BETWEEN THRESHOLD EXPERIMENT

Condition	N	M_1	M_{10}	σM_D	CR
30 mL—20 cps—Std. Int.	15	19.0	23.5	0.81	5.58**
30 mL—20 cps—3 min.	10	20.7	26.0	1.39	3.79*
30 mL—50 cps—5 min.	15	62.4	62.3	1.42	0.08
30 mL—20 cps—5 min.	21	21.0	22.0	0.72	1.38
30 mL—20 cps—10 min.	8	22.0	21.6	1.55	0.22

* Significant at .01 level of confidence.
** Significant at .001 level of confidence.

Experiment III concerns itself with the aftereffect on the phosphene threshold of 100 μA stimulation for three minutes. Table 5 shows the difference between prestimulation thresholds and thresholds at various intervals following the intense stimulation for the total 15 observers and a subgroup of five observers who were also run on the control condition. The curves in Figure 2 show these threshold differences for the total group and for the subgroup corrected for control values (line 4 in Table 5). It is seen that at five minutes following the intense stimulation the mean threshold for the group it still well above the initial level. The mechanism

TABLE 5
MEAN THRESHOLD DIFFERENCE BEFORE
AND AFTER PROLONGED STIMULATION

		Time in minutes						
	N	0	1	2	3	5	7	10
1. Total group	15	8.7	5.7	4.6	4.7	4.3		
2. Subgroup with additional intervals	5	6.1	2.5	3.7	3.4	2.2	1.5	2.1
3. Control	5	1.0	0.8	0.7	0.1	0.2		1.2
4. Difference 2 – 3	5	5.1	1.7	3.0	3.3	2.0		0.9

involved in the aftereffect seems to be very slow. Another interesting feature of this curve is that there is a slight increase in the curve at three minutes. Of the fifteen observers, nine showed an increased threshold at three minutes, two showed an increase at two minutes and two showed an increase at two as well as 5 minutes. Of two observers showing a continuous decrease, one had a levelling at two to three minutes. These individual differences in delay of threshold increase tend to obscure the point of deflection in the curve for the total group. With only one observer going continuously down, and the remaining fourteen showing either a rise or a plateau at one or another time interval, delayed decrease in sensitivity following prolonged stimulation seems to be demonstrated. It thus appears that there are two factors involved in the aftereffect of a prolonged and intense stimulus: one which causes the initial increase of the threshold immediately after intense prolonged stimulation, and another which comes into play only after some latency, and which does not reach its peak until from two to three minutes after cessations of the stimulation. It is indicated that the time course of these factors showed some individual differences.

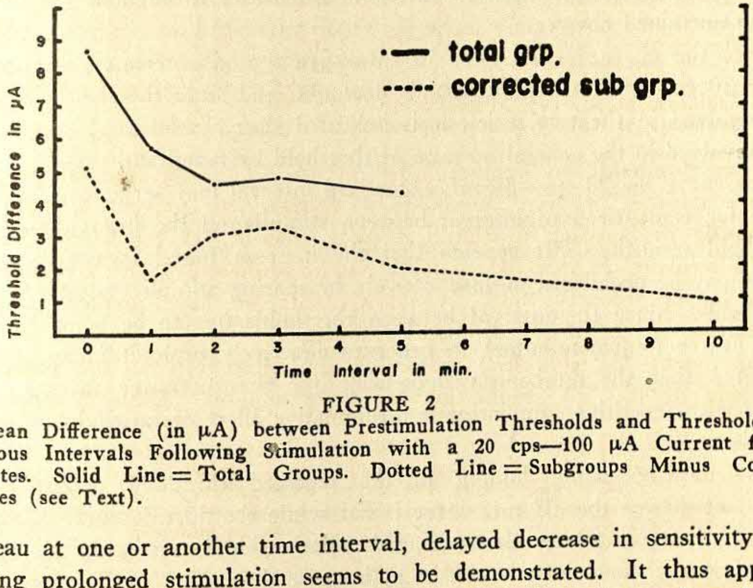


FIGURE 2

Mean Difference (in µA) between Prestimulation Thresholds and Thresholds at Various Intervals Following Stimulation with a 20 cps—100 µA Current for 3 minutes. Solid Line = Total Groups. Dotted Line = Subgroups Minus Control Values (see Text).

D. Discussion

Our findings show that electrical stimulation of the visual system results in a temporarily decreased excitability of the stimulated structure under certain conditions. We have demonstrated that this is the case whether the stimulation is four to five times threshold intensity or about threshold intensity.

We have seen from Figure 1 that the standard interval gives results practically identical to an interval of five seconds between stimuli. This allows a comparison of 20 cps—.03 mL between untrained observers (standard interval) and trained observers (five-second interval). While untrained observers have a gradually increasing threshold with five repetitions, the trained observers have a virtually flat threshold curve for ten repetitions. The difference between the two groups may be explained in terms of decreased attention to threshold intensities throughout the session by the untrained observers.

Since the curve for 20 cps—30 mL—five seconds increases consistently while 20 cps—.03 mL—five seconds does not, and since the same observers are involved in the two series, it is indicated that physiological factors are also involved in the general increase of threshold by repetition.

The curve for 20 cps—30 mL—standard interval may serve as a common point for comparison of interval between stimuli and the interval between threshold conditions. It appears that the increase found in the reference curve may be prohibited more effectively by spacing stimuli than by spacing thresholds. Since the interval between thresholds has to be comparatively long (three to five minutes) in order to effectively prohibit increase, it is indicated that the mechanism responsible for the successive threshold increase with repetitive stimulation is a summative effect of rather long lasting duration.

The most puzzling finding in the repeated threshold determination is that at 20 cps the .03 mL curve is flat while the 30 mL curve increases consistently. For 50 cps, however, this relationship is reversed. Here, the .03 mL curve shows a more consistent increase than the 30 mL curve. The fact that the 20 cps—30 mL condition results in a greater consistent increase than the 50 cps—30 mL condition shows that the threshold increase is not as directly related to magnitude of the threshold as we had expected.

The reversed relationship between the light situations for the two frequencies is not easily understood but is in agreement with the previously reported relationship between stimulus frequency and level of light adaptation (Clausen, 1955, 57 ff). Obviously the photic stimulation has a differential modifying influence on the phosphene threshold for 20 and 50 cps. It seems reasonable to believe that the discharge rate of the optic nerve resulting from photic stimulation is higher for the 30 mL than for the .03 mL condition, and that the difference in threshold for the two light conditions may be due to some kind of interaction between such discharge rate and the frequency of the electrical stimulation.

The increase for repeated threshold determination which occurs in some but not in other situations, is difficult to explain. An important point is that for each frequency it is the light condition giving the highest threshold which shows the most consistent increase. The rise in threshold with repetitive stimulation for the high threshold condition for each frequency must somehow be related to the fact that photic stimulation modifies in a differential manner phosphene threshold at 20 and 50 cps. An attempt to explain this in terms of discharge rate of the optic nerve resulting from photic stimulation in interaction with electrical stimulation, does not seem to be consistent with all the data. Since the rise in threshold seems to be a consequence of the modifying influence of photic stimulation on the threshold the most promising approach would seem to be further investigation of the mechanism involved in the influence of light on the phosphene threshold.

Of considerable interest are the results of the prolonged intense stimulation which demonstrates two factors operating in the aftereffect. It seems reasonable to assume that the first factor, which causes the initial increase after stimulation, is the same as that which causes the build up of consecutively recorded thresholds, since it begins to decline immediately and is obtained under the same conditions of light and frequency. It is, however, difficult to conceive of an interaction between neural events from the illumination and the imposed electrical frequency taking place two to three minutes after cessation of the electrical stimulus and lasting 10 minutes.

This unexpected second factor which reaches its peak two to three minutes after cessation of the stimulus and which has a remarkably slow decline, is even more difficult to explain than the first factor. The most obvious slow process in the visual system is dark adaptation. It might therefore be speculated that the 100 μA—three minutes stimulation acts in the same way a light adaptation, i.e., involving similar photochemical reactions. Following the electrical stimulation, then, one might have a slow process functionally equivalent to an adaptation to a higher illumination which terminates two to three minutes later whereupon a readaptation to the original illumination begins. Still the time course of this process is a puzzle since one would not expect an adaptation to a higher level taking place with such delay or such a slow readaptation to a still relatively bright illumination. It is, however, consistent with this explanation that at 20 cps the phosphene threshold increases with adaptation to increasing light intensities (Clausen, 1955, 57 ff).

If our interpretation for the delayed aftereffect is correct it would mean that the electrical stimulation acts directly on the photochemical processes in a surprisingly slow manner, reaching a maximum effect two to three minutes after cessation.

Since phosphene thresholds at frequencies from 40 to 70 cps have a reversed relationship to adaptation level, evidence for this hypothesis might be obtained by using 70 cps with the expectation that the threshold would now temporarily decrease rather than increase at two to three minutes following stimulation. The 10-minute duration of the aftereffect could be tested by recording phosphene threshold when adapted to a bright light, substitute the prolonged electrical stimulation by a brighter photic stimulus, and again record comparison thresholds at various intervals following this stimulation. If 20 cps is used for the phosphene threshold determination, an increase in threshold should follow. If a high frequency is used, the result would be a lowered threshold.

According to our explanation, the 20 cps—.03 mL condition would be most sensitive to the delayed aftereffect. If this be so, the fact that with repeated threshold determinations no increase occurs for this condition is consistent with our notion that different mechanisms are involved in prolonged stimulation and repeated threshold determinations.

A further possibility is that the electrical stimulation is affecting the chemical climate of the synapses in such a way that a condition inhibitory to transmission is created. It is still strange, however, that such a delay occurs. It would be interesting to see if similar results are obtained after prolonged electrical stimulation to another modality. If such are obtained, a more general process than the photochemical process would be indicated.

It is difficult to generalize from our data to that of Riggs, Cornsweet and Lewis (1957) since their conditions are quite different, i.e., the thresholds were obtained on the dark adapted eye using D. C. stimuli of 0.1-second duration with 7.65 or 15.3 seconds between stimuli. It is pertinent to point out, however, that their data do seem to demonstrate an increase in threshold when the Constant Stimuli method is used. This increase is quite compelling if one keeps in mind the pulse duration and interval used as well as the fact that a three-minute interval intervened between the two determinations. On the other hand, thresholds determined by their Limits-Comparison method do not seem to show a comparable consistent increase. They also seem to have demonstrated some individual difference in decreasing sensitivity with repeated stimulation.

Thus on the basis of our interpretation of the Riggs *et al.* data and the results of our present study we are inclined to agree with Kohata, Komatsu, and Motokawa (1956) that repetitive electrical stimulation of the eye can lead to decreased sensitivity. We are, however, not in a position to evaluate the extremely long-lasting decreased sensitivity (up to an hour) reported by them.

Heretofore the action of electrical stimulation of the eye has been customarily discussed in terms of electrophysiological processes of the stimulated structures. Our results have led us to speculate that an electrical current may initiate a chemical process or influence an existing one.

E. SUMMARY

In an effort to investigate aftereffect of electrical stimulation of the eye on phosphene threshold we analyzed data from a previous study utilizing untrained observers and data from three especially designed experiments where we (*a*) varied the time interval between each stimulus; (*b*) varied the time interval between each successive threshold determination; and (*c*) compared the threshold before a three-minute electrical stimulation to a threshold obtained at varying intervals after this stimulation.

It was found that: (*a*) The threshold increases significantly after five consecutive threshold determinations for untrained observers but does not increase for trained observers under the same conditions. (*b*) In trained observers systematic threshold increase occurs for some combinations of stimulus frequencies and light conditions but not for others. (*c*) Where increases occur they may be counteracted by spacing stimulus presentation or threshold determination. (*d*) Immediately after prolonged intense stimulation there is a threshold increase which declines for the first two to three minutes. At this point there is a second increase with a return to original threshold level in about 10 minutes.

From the data we have concluded that attention is responsible for the difference between trained and untrained observers. It has been speculated that there are two physiological mechanisms producing aftereffects. One causing the immediate threshold rise after prolonged as well as repetitive stimulation, and which seems to be an interaction between stimulus frequency and light condition. The other, characterized by an initial delay, may be a result of the electrical stimulation affecting chemical processes.

REFERENCES

1. CLAUSEN, J. Visual Sensations (Phosphenes) Produced by AC Sine Wave Stimulation. Copenhagen: Ejnar Munksgaard, 1955.
2. CLAUSEN, J. & KARRER, R. Phosphene threshold as related to age and sex. *J. of Psychol.*, 1959, **47**, 189-198.

3. Kohata, T., Komatsu, M., & Motokawa, K. Aftereffects of electrical test shocks as a factor affecting retinal color effects. *Japan. J. Physiol.*, 1956, **6**, 236-248.

4. Riggs, L. A., Cornsweet, J. C., & Lewis, W. G. Effects of light on electrical excitation of the human eye. *Psychol. Monog.*, 1957, **71**, #5.

The Training School
Vineland, New Jersey

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 119-125.

STUDENTS' ATTITUDES TOWARD COMPULSORY PARTICIPATION IN EXPERIMENTS*

Department of Psychology, New York University

ALICE GUSTAV

A. INTRODUCTION

The famous statement made several decades ago that American psychology may be considered the psychology of the American college sophomore is still largely true because undergraduate psychology classes continue to be a readily available source of subjects. Experimental reports customarily include a statement warning the reader that conclusions must not be generalized beyond this population and that formulation of experimental design has taken account of whether the subjects are naive as to the purpose of the experiment. But little attention has been paid to the crucial question of students' feelings about being subjects. Brower (1) did find significant differences in mean scores on a visuo-motor test between voluntary and compulsory groups of college students, which was an indication that experimental results may well be affected by such attitudes. The current study reports an attempt to determine what college students felt about compulsory participation in experiments.

B. SUBJECTS AND PROCEDURE

Subjects in this study were the students registered in all introductory psychology classes in Washington Square College (an undergraduate liberal arts division of New York University) in the 1961 spring term. By departmental rule, each is required to serve as a subject in one current professional experiment. Most instructors made an appropriate announcement at the beginning of the term to pave the way for research people coming in later to arrange appointments. The latter were primarily doctoral candidates and persons working on research grants.

During one session in the last week of the term this investigator, personally or by proxy, visited each class and asked the students to fill out, anonymously, the sentence completion test shown in Table 1 which had been constructed for this purpose. The advantages of this technique have been reported (2) as well as its use for student evaluation of course work (3).

---

* Received in the Editorial Office on September 13, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

A total of 281 replies were obtained. This represented all students who had not withdrawn from the course and who were present on the day the forms were filled out. Twenty-one students (16 of them boys) had simply written a statement that they had not acted as subjects. Obviously, the requirement was not completely policed and students who were determined to resist managed to evade performance. Another nine students (eight of them boys) were facetious, and can only be described as utterly antagonistic to the whole procedure. While the last mentioned 30 papers are excluded from the following evaluations, leaving 251 papers to be considered, they might be kept in mind when replies unfavorable to the program are discussed below.

TABLE 1
SENTENCE COMPLETION TEST

Age:

Sex:

Directions: Complete the sentences.

1. When the teacher announced that each student would have to be a subject in an experiment
2. When I came to the room in which the experiment was to be done
3. During the experiment
4. After the experiment was over
5. If participation in the experiment had been voluntary, instead of required,

Comments: (Add anything you wish to say)

C. RESULTS

When the responses were read it was found that they could easily be grouped into categories which indicated the students were: (*a*) favorably inclined to the whole process; or (*b*) unfavorably disposed, or (*c*) simply factual. Therefore, these are the broad classifications used below. Except for the 30 papers mentioned in the preceding section, responses for girls and boys were so similar that they have been combined.

It will be noted that totals for each category are always less than 251. This occurred because some students skipped an item or two. Also, in one class the instructor had not made an announcement of the requirement, but proceeded to use the subjects himself in a group study without advance notice. For these students Stems 1 and 2 were not pertinent and had to be omitted.

Table 2 contains the summary classification of responses to all the stems (stimulus phrases). In the presentation of results for each stem, to be given below, the parenthetical sentences are verbatim student responses which were typical of answers in that grouping.

TABLE 2

SUMMARY OF RESPONSES TO SENTENCE COMPLETION TEST

Stem No.	Favorable N	Favorable %	Unfavorable N	Unfavorable %	Factual N	Factual %	Total N	Total %
1	135	62	78	36	4	2	217	100
2	28	14	84	41	93	45	205	100
3	111	47	89	37	38	16	238	100
4	169	69	19	8	56	23	244	100
5	141	63	84	37	0	0	225	100

1. *Stem 1*

The 135 favorable responses broke down into three sub-categories: 54 indicated mild willingness to conform to a course requirement ("It seemed like a small request for a course.") ; 34 expressed curiosity about psychological experiments ("I wondered what an experiment in psychology would be like.") ; and 47 were eager to sample a new experience ("I felt that it would be a rewarding experience as I normally would not have had such an opportunity to participate in anything of this nature.").

Of the 78 unfavorable responses, 53 expressed annoyance ("I was annoyed at being a 'guinea pig' for some experimenter.") ; eight were fearful of the situation or their own inadequacy ("I pictured myself in a dark room with people giving me electric shocks." "I was scared I would fail.") ; and 17 felt the pressure of time ("I thought to myself of its interference with things I wished to do at the same time.").

The four factual responses were statements indicating what happened objectively, not revealing feelings of any kind, e.g., "The experimenter was ready to make appointments."

2. *Stem 2*

The 28 favorable responses were all expressions of interest and curiosity about the experiment ("I was interested in the workings of the apparatus." "My resignation developed into curiosity.").

The 84 unfavorable responses included 16 showing boredom or irritation ("I thought it was a nuisance.") and 68 of an apprehensive nature ("I was nervous." "I was anxious to get it over with.").

The 93 factual responses were, an indicated for Stem 1, strictly objective observations.

3. *Stem 3*

The 111 favorable responses were divided as follows: 31 said they were trying to do their best ("I tried hard to do it well.") ; 26 were interested ("I quite forgot to smoke, a sign of interest, if nothing else.") ; 28 were curious about the experiment ("I lost track of the count of the cards, I was so busy trying to figure out what this experiment would prove.") ; and 26 felt comfortable ("I felt at ease.").

Of the 89 unfavorable responses, 42 were just bored ("I participated as directed without enthusiasm.") ; and 47 were nervous, usually because of concern about their own performance ("I felt tense.").

Two of the 38 factual responses might be of interest to experimenters: "I copied from my friend's paper." and "The tester spyed (sic) on me." Whether these should be included among the "unfavorable" responses is a moot point.

4. *Stem 4*

The 169 favorable responses included 17 who expressed interest and enjoyment ("I thought it was fun.") ; 77 who experienced a relief from tension ("I was relieved—so relieved in fact, that I allowed her to talk me into taking another one.") ; 14 curious to know their own results; ("I wanted to find out the results of my test."), and 61 who were curious about the experiment ("I would have liked to know what was being tested and why.").

Of the 19 unfavorable responses, seven were disappointed ("I was a little disappointed because it did not live up to my expectations.") ; and 12 were bored ("I felt there was little motivation to do well.").

The 56 factual responses were quite routine.

5. *Stem 5*

Of the 141 favorable responses, 87 indicated they definitely would have participated anyway and 54 said probably they would have. Of the 87 who were definite, 73 gave no reason ("I still would have participated.") ; 12 said it would have been due to curiosity ("I would have gone anyway just for the experience.") ; and 2 wished to impress the instructor ("I still would have gone, hoping to make a favorable impression on the instructor."). Of the 54 who thought they might have done so, 20 gave no reasons ("I probably would have participated anyway.") ; and the remaining 34 gave a wide variety of reasons, e.g., curiosity, time, type of experiment ("I probably would have waited for an experiment that I would consider more interesting."), etc.

Of the 84 unfavorable responses, 77 were flat statements that they would not have done it ("I wouldn't have gone." "I would have followed the old army adage, never volunteer for anything!"); and seven said it was because of lack of time ("I would not have taken the time to participate.").

6. *Comments*

Only 37 students made spontaneous comments. Twelve favorable ones reiterated interest and enjoyment. The 25 unfavorable comments included eight concerning pressure of time, four criticizing the procedure because there had been no discussion of how the experiments tied in with class work, nine expressing derogatory feelings about content and method of psychology in general, and four largely irrelevant to this study.

D. DISCUSSION

Students' attitudes toward being subjects may have important repercussions in both experimental results and in the public image of psychology.

The findings of this study suggest three possible ways in which experimental results might be distorted as a result of subjects' attitudes. One such situation is that in which subjects perform in a lackadaisical or perfunctory fashion. Students who reported they were bored or felt pressed for time may well have been in this group. That is, they would go through the motions in order to be released quickly, without exerting themselves unduly.

A second source of possible distortion would come from those subjects, small in number, who might deliberately twist their responses as revenge for what they consider an imposition. Here we have the few students who said they were irritated because it was compulsory, and we might include here the nine who wrote facetious statements on the sentence completion form. Incidentally, this would seem to be more of a problem with boys than with girls. As noted earlier, only five girls evaded the requirement completely and only one handed in a facetious paper, which confirms the general finding of many other studies that girls tend to be more obedient to school authority.

The third factor which might affect experimental results adversely, and definitely the largest in terms of numbers of students involved, is apprehension of what would happen during the experiment and a fear of being measured and found wanting. This becomes apparent if we follow the sequence of responses. Only eight admitted to fear early in the term when informed they would have to be subjects (Stem 1). However, when they found themselves about to face the situation (Stem 2) 68 admitted apprehension and the smallest number of favorable responses were given. Many of the people who marked Stem 1 favorably had shifted to "factual responses" at Stem 2 which may indicate at best acceptance, but certainly not enthusiasm. While the experiments were in progress (Stem 3) 47 said they continued to be nervous. When the experiments were over (Stem 4) 77 reported relief from tension, and at this point there were the smallest number of unfavorable and the largest number of favorable answers. A little tension may act as a motivating force in a subject's performance, but too much could be paralyzing, thus changing experimental results.

There are also implications in this study for public relations. Generally, clinical and industrial psychologists have been most active in the struggle to achieve a positive and informed public awareness of psychology. Results in this study would indicate that there is a powerful role which academic psychologists, teachers and experimentalists, can play in molding the attitudes of our future public, i.e., undergraduate students, many of whom take no more than the introductory course. An examination of the results shows that these students, at one time or another, were eager to learn what the psychologist was doing and why. Themes which carried through all the stems were "curiosity" and "interest," despite apprehension. Although they are fearful, they want to learn about the methods of psychology. When first told about the need to be subjects (Stem 1) 34 were curious about psychological experiments and 47 were eager to have a new experience; when faced with the actual task (Stem 2) the entire total of 28 favorable responses was in terms of interest and curiosity; during the experiment (Stem 3) 26 were interested and 28 curious about the experiment; and when it was over (Stem 4) 17 were interested, 14 curious about their own results and 61 curious about the experiment. And, as noted above, when they have just completed participation in an experiment is when we find the nadir of unfavorable, and the peak of favorable, student attitudes. Obviously, that is the advantageous place for the experimenter to take a little time to make the student feel he is not simply a "guinea pig," and to assuage his intellectual curiosity in a manner conducive to continued interest. This might take the form of a brief explanation of the purpose of the experiment, a promise (carried out) to give individual results at a later time, or brief mention as to how the topic of the experiment relates to material in the student's textbook. In addition, the classroom teacher might ascertain from the experimenter something of what is being done, which could then be transmitted to the students during lectures. Admittedly, this takes time. But the effort may bring great future dividends to the field of psychology.

E. Summary

Attitudes about compulsory participation in experiments were elicited from undergraduate students by means of a sentence completion test. Their answers suggested that experimental findings may be distorted as a result of subjects' apathy, deliberate change, or apprehension. There were also suggestions that academic psychologists may aid in promoting favorable attitudes on the part of students toward psychological experimentation.

References

1. Brower, D. The role of incentive in psychological research. *J. Gen. Psychol.*, 1948, **39**, 145-147.
2. Gustav, A. Evaluation of dentist-patient relationships by means of a psychological test. *In* J. L. Blass, (*Ed.*) *Motivating Patients for More Effective Dental Service.* Philadelphia: Lippincott, 1958. Pp. 5-11.
3. Lindgren, H. C. The incomplete sentences test as a means of course evaluation. *Educ. Psychol. Measmt.*, 1952, **12**, 217-225.

Department of Psychology
New York University
New York 3, New York

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 127-142.

AN EMPIRICAL THEORY OF THE SERIAL-POSITION EFFECT*

Department of Education, University of California

ARTHUR R. JENSEN

A. INTRODUCTION

In rote-learning a list of nonsense syllables or other relatively homogeneous items, the errors made before mastery are distributed according to the positions of the items in the series, forming the classical, negatively skewed, bow-shaped serial-position curve. The problem of explaining this curve has long occupied learning theorists, but no satisfactory theory has yet emerged.

The facts and theories of serial learning have been reviewed by McGeoch and Irion (9). Some of the important facts that must be accounted for by a theory of serial learning are the following: (*a*) factors affecting the overall difficulty of the learning task, such as familiarity of the items and distribution of practice, do not much affect the *relative* difficulty at each position, i.e., all serial-position curves for a list of a given length are extremely similar when plotted in terms of the *percentage* of total errors at each position (8); (*b*) individual differences in learning ability and factors (e.g., anxiety) which affect learning rate and memory span, affect the degree of skewness more than the degree of bowing of the serial-position curve (9, pp. 124-125); (*c*) the degree of skewness decreases as the number of items in the list increases; (*d*) a continuous list in which the temporal separation between the last item and the first item is the same as the separation between other items in the list produces a flatter bow-shaped serial-position curve than when there is a temporal gap between presentations of the list (3).

1. *An Empirical Theory*

The purpose of this paper is to present an empirical theory of the serial-position effect, along with some supporting evidence. The theory is called "empirical" because it is derived largely from experimental findings and is not based on assumptions or postulates having little or no empirical

---

* Received in the Editorial Office on September 13, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

foundation. It is not offered as a highly formal theory or as a final finished product, but as an heuristic hypothesis that opens the way for further investigation.

The shape of the serial-position curve would seem to be a consequence of the following conditions:

a. The number of items or responses to be learned in the serial list is greater than the subject's (*S*'s) immediate memory span, which is the length of list he can learn in one trial. Consequently, the *S* requires more than one trial or exposure in order to learn the list. Without this condition, there can be no serial-position curve.

b. If the *S* cannot learn all the items on one trial, he must learn some items before he learns others; this implies that the items must be learned in in some *order*.

c. The first item learned is usually the first item to which the *S* attends or to which the *S* makes some response. This item is usually the first one presented in the serial list. It may be regarded as the *anchor* item.

d. If it is assumed that the *S* learns most readily by "attaching" new responses to previously learned responses, the *S* should learn the serial list around the anchor item, both in a forward and in a backward direction. The anchor serves as a cue for other responses in the series. The basic assumption is that it is easier to learn an unfamiliar item that is adjacent to an item that has already been learned than it is to learn an unfamiliar item that is presented among other items not yet learned.

e. If this is true, the serial list should be learned (when learning is by the usual anticipation method) in a certain order around the anchor point. Take, for example, a 9-item serial list. Say our hypothetical *S* learns one item on every trial; the first item is established as the anchor position on the first trial. On the second trial he will learn Item 2, on the third trial Item 9, and so on. Thus, the order of learning will be around the anchor point as follows:

Serial position: 6 7 8 9 $\underline{\underline{1}}$ 2 3 4 5
Order of learning: 9 7 5 3 $\underline{\underline{1}}$ 2 4 6 8

Or:

Serial position: 1 2 3 4 5 6 7 8 9
Order of learning: 1 2 4 6 8 9 7 5 3

f. An additional assumption, which is already well-supported by empirical evidence, applies to serial lists of fewer than seven or eight items. For such short lists, the first two or three items are learned in a single trial. This is probably due to the fact that in a short list, which is just three or four items longer than the *S*'s immediate memory span, there is not so much retroactive interference from the last two or three items in the list to "inhibit" the *S*'s retention of the first few items which fall within the *S*'s memory span. Thus, for a 7-item list, the order of learning would be:

Serial position:	1	2	3	4	5	6	7
Order of learning:	1	2	3	5	7	6	4

And for a 6-item list the order would be:

Serial position:	1	2	3	4	5	6
Order of learning:	1	2	3	5	6	4

Consequently, the shorter the list, the more skewed the serial-position curve. The curves for long serial lists would appear almost symmetrical.

g. If it is assumed that the number of trials required by the *S* to learn the serial list is equally divided among the items in the series, and if each item is learned on one trial in an all-or-none fashion, the proportion of the total errors made at each position up to the trial of mastery of the whole list should be equal to the number of the rank order in which that position is learned divided by $1 + 2 \ldots + N$, where N is the number of positions in the series. For example, the percentage of total errors made at each position in a 10-item series learned to a criterion of mastery would be:

Serial Position:	1	2	3	4	5	6	7	8	9	10
Order of learning:	1	2	4	6	8	10	9	7	5	3
Percentage errors:	1.82	3.64	7.27	10.91	14.55	18.18	16.36	12.73	9.09	5.45

The average serial-position curve for a *group* of *S*s is usually flatter and more bow-shaped than the almost triangular-shaped curve predicted by the theory. The reason for this is that the group curve actually does not represent the curve for individual *S*s. Since there is not perfect unanimity among *S*s in the order in which items are learned, though the degree of agreement is indeed high, the group curve appears to be a "smoothed" version of the curve predicted by the theory. The theory is "probabilistic" in the sense that it only predicts the most probable order of learning the items. Because of the many idiosyncratic factors that affect a given *S*'s performance, there are slight variations in the order of learning; certain items will tend, because of idiosyncratic associations, to become secondary anchor points in the series. And for lists that are learned in just a few trials, the characteristic oscillations in performance from trial to trial can considerably affect the degree of irregularity of those aspects of the data from which we determine the rank order in which the *S* learned the items, viz., the number of errors at each position or the number of trials preceding a given criterion of learning for each position. The theory, therefore, predicts that differences in the degree of bowing of group serial-position curves are caused by differences among the groups in the degree of within-group agreement in the order of learning. Thus, a relatively flat serial-position curve would result from relatively low agreement among *S*s in the order of learning, while a very peaked curve would indicate high agreement among *S*s in the order of learning the items. Differences in skewness, on the other hand, would be related to conditions which affect memory span.

B. METHOD

1. *Subjects*

The *S*s were 245 university students recruited from an introductory course in educational psychology. There were approximately twice as many women as men.

2. *Procedure*

Learning a list of nonsense syllables or other verbal materials generally involves two phases: (*a*) the acquisition of the syllables; and (*b*) the learning of their serial order. In order to minimize the first phase, since we wished to study mainly the *serial* learning, the present experiments used as stimuli colored geometric forms: triangles, squares and circles colored red, yellow, blue, and white. Nine- and ten-item lists were composed of these stimuli. Two rules governed the formation of lists: (*a*) no item would appear more than once in the list; and (*b*) items of the same shape or color would never be adjacent in the series. Before the serial learning task was presented, the *S* always knew perfectly the items composing the list. The stimuli were projected one at a time on a screen at rates varying in different experiments from two to four seconds per stimulus, with intertrial intervals of from four to eight seconds. A technical description of the apparatus is presented elsewhere (6). *S*s learned by the anticipation method, responding by saying "red triangle," etc. They were encouraged to guess when in doubt as to the correct response on subsequent trials. *S*s learned to a criterion of mastery of the whole list.

C. RESULTS

Only those aspects of the data directly relevant to the proposed theory of the serial-position effect are presented.

1. *Evidence of the One-Trial Nature of Serial Learning*

A backward learning curve (4), showing the twelve trials preceding the criterion of mastery was plotted for the last position learned in the series by each of 200 *S*s. The last item learned was defined as the item on which the last error was made before mastery of the whole list; in the case of ties, the item closest to the middle of the series was selected. The backward "curve" is shown in Figure 1, which clearly indicates that there was no significant rise above a chance level in the correct responses among the 200 *S*s on the several trials before the criterion was attained. Backward curves plotted in the same manner for each position in the series, with criterion of learning for each position being three successive correct responses, are essentially identical to Figure 1. To determine the chance percentage of correct responses that could be obtained by sheer guessing, 20 *S*s were forced to guess for 20 trials, with a different serial order on every trial. The mean percentage correct by sheer guessing was 28.30 per cent. These results indicate a sudden, rather than a gradual, learning of the serial position of the individual items.

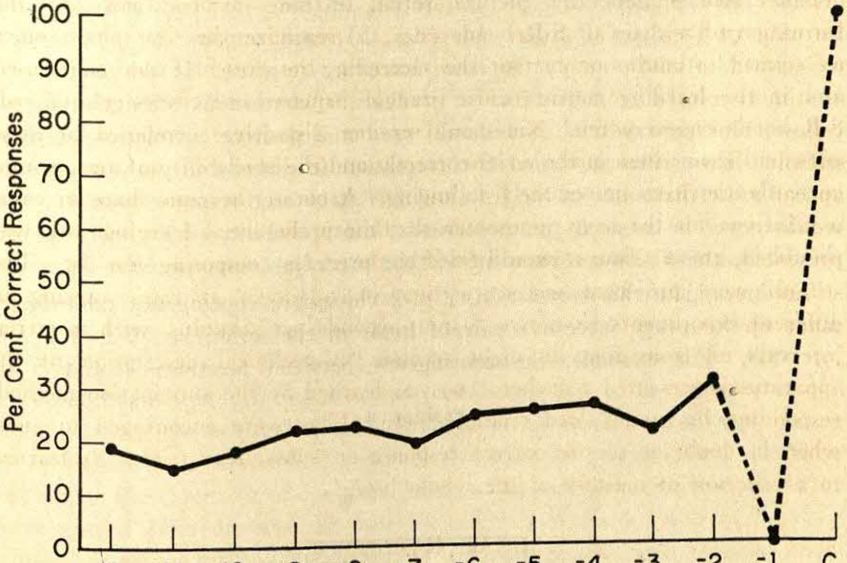


FIGURE 1
A backward learning curve for the last item learned, showing the percentage of correct responses among 200 *S*s on each of the 12 trials preceding the criterion of mastery.

The percentage of errors made at each position up to the trial at which the item had been anticipated correctly on three successive trials is at least as great as the percentage of errors obtained under conditions of chance guessing. For 60 *S*s who learned a nine-item list at a four-second rate with an eight-second intertrial interval, the percentage of errors at each position up to the criterion trial for that position was:

Position:	1	2	3	4	5	6	7	8	9
Per cent errors:	81	79	84	78	82	83	85	82	81

The mean percentage of errors was 81.74. The mean percentage of errors on the very first trial, which represented sheer guessing, was 79.75 per cent. Thus, these results substantiate the impression that no learning was manifest at a given position up to the trial on which the item in that position attained the criterion of learning. The rather stringent criterion of three successive correct responses was needed to distinguish *learned* correct responses from correct responses that might be due to chance.

2. *The S-R Contingencies in Serial Learning*

S-R theories generally picture serial learning as consisting of the formation of a chain of S-R bonds (e.g., 5), each response in turn serving as a partial stimulus or cue for the succeeding response. If this were true, and if the learning consists of a gradual or continuous strengthening of S-R bonds on every trial, one should predict a positive correlation between anticipating one item in the series correctly and the probability of anticipating correctly the item immediately following. A correct response would act as a valid cue for the next response so that the probability of its being correct should be greater than it would be if the preceding response were incorrect.

To investigate this correlation, 2 × 2 chi-square contingency tables were made up for every successive pair of items in the series, i.e., 1-2, 2-3, 3-4, . . . 8-9. For example, the contingency between positions 4 and 5 is:

		Position 4	
		−	+
Position 5	+	5	8
	−	14	23

$\chi^2 = .002$

The cell entries were based on the first trial on which each of 60 *S*s made four correct responses on a 9-item list. (Only 50 *S*s in this group had a trial with exactly four correct responses.) Another set of contingency tables was based on the first trial on which each *S* made five correct responses. (Only 54 *S*s had a trial with exactly five correct responses.) There were 16 contingency tables in all. Of these 16 contingencies, only four yielded chi-squares large enough to be significant at the 5 per cent level, and of these only three were in the predicted direction. The highest degree of correlation for any of the contingencies is represented by a phi coefficient of .16. It may thus be concluded that the correlation between successive serial responses on any given trial is negligible.

This fact seems surprising in view of the traditional S-R theories of serial learning, which assume an incremental, rather than an all-or-none, growth in associative strength between items. The interesting and highly important point, however, is that these findings are perfectly consistent with a one-trial, all-or-none conception of learning. If the *S* responds at a strictly chance level at each position until on one trial the item in a particular position is learned, there should be a zero correlation on any one trial between whether or not a response is correct and whether or not the immediately succeeding response is correct. This actually seems to be case.

3. *The Relative Difficulty of Learning Each Position*

The percentage of total errors that occurred at each serial position during the course of learning a 9-item list (stimuli presented at a two-second rate with a four-second intertrial interval) up to a criterion of three successive perfect repetitions of the whole list was determined for each *S* ($N = 60$). These percentages for each *S* were arranged in the order in which the corresponding items were learned to a criterion of three successive correct trials; the percentages were then averaged for the entire group of 60 *S*s. As can be seen in Figure 2, the increment in percentage of errors for each successive item learned is a *constant proportion* of the total errors made in learning the entire list. An analysis of variance showed that the points in Figure 2 do not depart significantly from the straight line, which was drawn to the closest fit of the data points. The same straight line would result if instead of errors we had used the number of trials taken to learn each item. The correlation (Pearson *r*) between errors and number of trials to criterion was .92. If we assume that items are learned in an all-or-none fashion, Figure 2 may be interpreted as indicating that, for a given *S*, every item in the series takes as many trials to be learned as every other item, once the previous items in the order of learning have been learned.

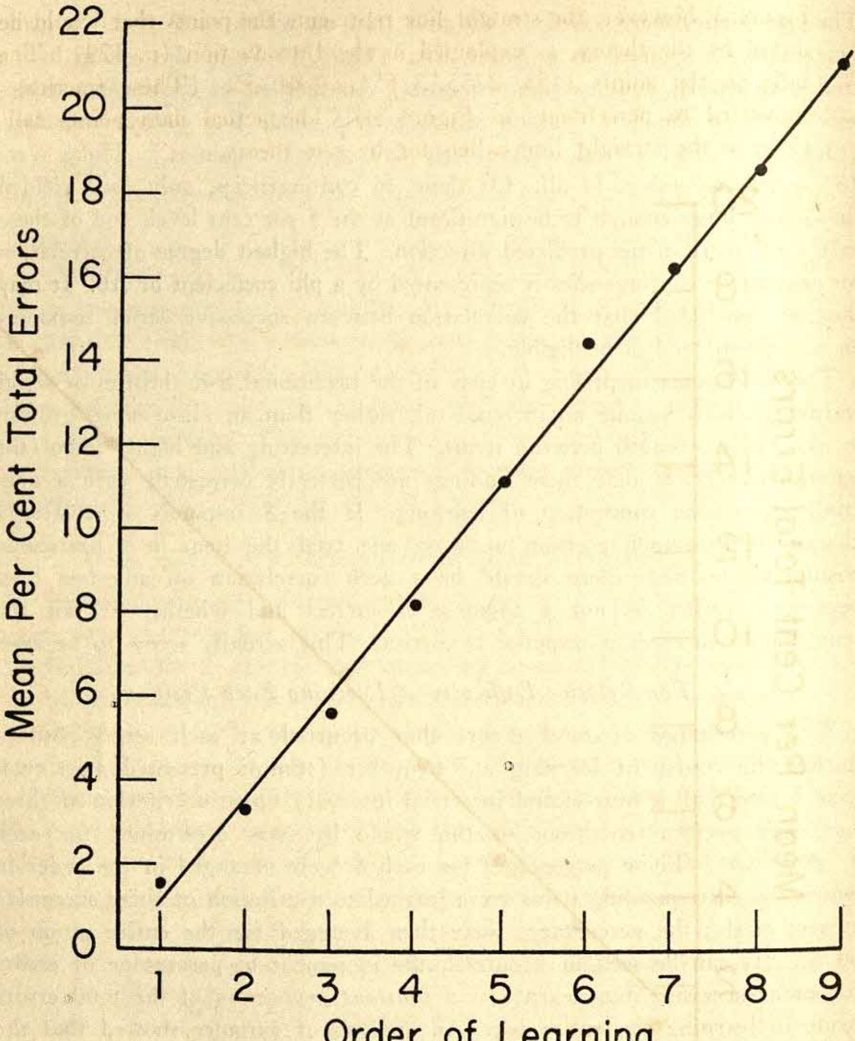


FIGURE 2
Mean percentage of errors on each item plotted in the order (determined for each *S*) in which each item in the 9-item serial list was learned to a criterion of three successive correct trials. ($N = 60$). The straight line is drawn to the best fit.

Figure 3 was obtained in the same manner as Figure 2; it is based on 25 *S*s who had learned a 10-item list to a criterion of mastery, with a three-second rate of presentation and a six-second intertrial interval. In

Figure 3, however, the straight line represents the points that would be predicted by the theory, as explained in the Introduction (p. 129). The line falls on the points 1/55, 2/55, 3/55 . . . 10/55. (These fractions were converted to percentages in Figure 3.) The actual data points fall very close to the straight line called for by our theory.

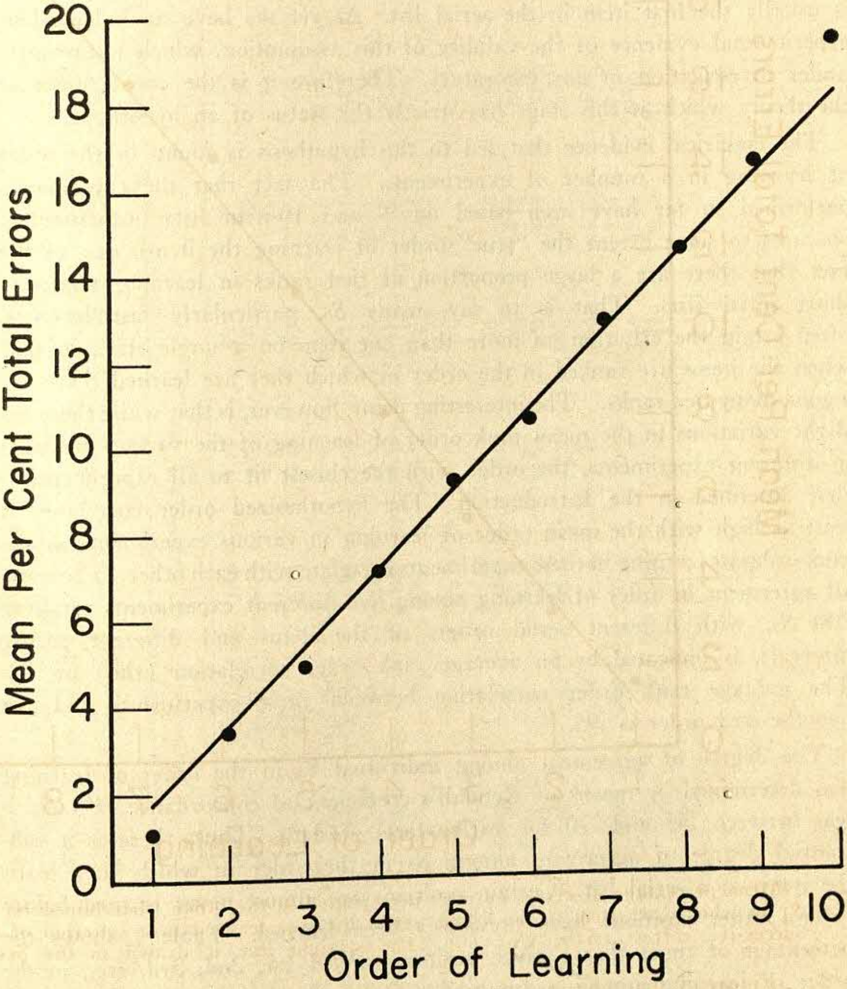


FIGURE 3

Mean percentage of errors on each item plotted in the order (determined for each *S*) in which each item in the 10-item serial list was learned to a criterion of mastery (one perfect trial). ($N = 25$). The straight line is called for by the theory.

4. *Order of Learning*

The hypothesis concerning the order of learning is based on theoretical and empirical considerations. It is held to be easier to learn new responses when they are adjacent to known, established, or anchored responses, and therefore the *S* learns the serial list around the first-learned item, which is usually the first item in the serial list. As yet we have no independent experimental evidence of the validity of this assumption, which is presently under investigation in our laboratory. Therefore it is the one element of the theory which at this stage has strictly the status of an hypothesis.

The empirical evidence that led to this hypothesis is found in the order of learning in a number of experiments. The fact that the experiments performed so far have been based on 9- and 10-item lists unfortunately obscures to some extent the "true" order of learning the items, due to the fact that there are a large proportion of tied ranks in learning relatively short serial lists. That is to say, many *S*s, particularly fast learners, often attain the criterion on more than one item on a single trial, so that when the items are ranked in the order in which they are learned there are a good many tied ranks. The interesting point, however, is that while there are slight variations in the mean rank order of learning of the various positions in different experiments, the order with the closest fit to all experiments is that described in the Introduction. The hypothesized order correlates at least as high with the mean order of learning in various experiments as the rank order of learning in these experiments correlate with each other The overall agreement in order of learning among five different experiments totalling 200 *S*s, with different serial orders of the items and different pacing intervals, is indicated by an average rank order correlation (rho) of .92. The average rank order correlation between these experiments and the hypothesized order is .95.

The degree of agreement among individual *S*s in the order of learning was determined by means of Kendall's coefficient of concordance *W*, which was between .55 and .70 for various sets of data. Thus, there is a substantial degree of agreement among *S*s in the order in which they learn the items in a serial list. Certain positions are almost never learned before certain other positions have previously been learned. Table 1 shows the percentage of times that a given position ranked 1st, 2nd, 3rd, etc., in the order of learning among a group of 120 *S*s who learned a 9-item list to mastery. The tied ranks, which were tabulated separately, are shown in the lower half of the table.

TABLE 1

Percentage of Occurrence of Each Rank Order of Learning at Each Position for 120 *S*s

Rank order	Serial position 1	2	3	4	5	6	7	8	9
1	53.49	27.91	3.49	1.16	.00	.00	.00	.00	13.95
2	28.37	31.07	16.21	4.05	.00	.00	1.35	.00	18.91
3	4.00	24.00	34.66	5.33	.00	1.33	1.33	4.00	25.33
4	1.70	13.56	23.73	30.51	.00	5.08	.00	10.17	15.25
5	4.70	2.35	16.46	22.34	9.41	9.41	2.35	14.11	18.82
6	1.43	1.43	2.86	12.85	15.71	15.71	20.00	28.56	1.43
7	1.11	1.11	4.44	15.55	15.55	21.11	20.00	21.11	.00
8	1.59	.00	1.59	9.52	30.15	26.98	19.04	11.11	.00
9	.00	.00	1.28	2.56	34.61	17.95	35.90	7.69	.00
1.5	40.82	30.62	8.16	4.08	.00	.00	2.04	.00	14.29
2.5	19.57	28.26	10.87	6.52	.00	4.35	2.17	2.17	26.09
3.5	8.89	11.11	26.67	11.11	.00	4.44	4.44	4.44	28.89
4.5	2.50	5.00	20.00	22.50	5.00	7.50	2.50	10.00	25.00
5.5	.00	3.45	10.34	24.14	3.45	13.79	10.34	31.03	3.45
6.5	2.17	6.52	8.70	19.57	8.70	13.04	15.22	26.09	.00
7.5	.00	.00	3.13	9.37	15.63	28.13	25.00	18.75	.00
8.5	.00	.00	3.17	3.17	30.15	25.39	25.39	11.11	1.59

5. *Order of Learning Serial Lists of Verbal Materials*

How well does the hypothesized order correspond to the rank order of errors in serial-position curves reported in the literature, which are based on lists of nonsense syllables and other verbal materials? To find out, the investigator collected all the serial-position curves he could find in the experimental literature—70 in all. They were based on lists of from six to 18 items. The mean rank order of learning the items as inferred from the data points in the serial-position curves was determined for every length of list. The Pearson *r* was used as the measure of correlation between the actual order of learning found in the literature and the theoretical order. Table 2 presents the results. It should be noted that the correlations, representing the "fit" of the theory to the actual data, are approximately as high as the "reliability" of the rank order of the actual data. The coefficient of concordance, *W*, is a measure of the average agreement between the individual lists. It is generally over .90. A number of other possible rank orders were correlated with the actual data to see if any better fit could be obtained, but no other order has as consistently close a fit, when measured by Pearson *r*, as the order described in the theory section of the Introduction.

TABLE 2

Correlations (Pearson *r*) Between Rank Order of Learning Items in Serial Lists Reported in the Literature and the Theoretical Order of Learning

Number of positions in serial-position curve	Number of curves	W*	*r*
6	2	.91	.97
8	9	.93	.98
9	7	.91	.96
10	21	.73	.97
11	1	—	.99
12	16	.85	.98
13	2	.95	.96
14	7	.92	.97
15	4	.98	.93
18	1	—	.86

* *W*, the coefficient of concordance, indicates the degree of agreement between the individual curves.

D. Discussion

Supportive evidence for the proposed theory is found in a number of highly relevant experiments in the literature.

The one-trial, all-or-none conception of serial learning is given support in an experiment by Bolles (2), who changed the items in the middle positions of a serial list one-third of the way through learning and found that the number of trials to learn the middle positions did not differ significantly from the number required by a control group for whom there was no change in the serial list throughout the course of learning. Bolles concluded, "The results of this study support the view that the associative strength of S-R sequences in the middle of the list is not gradually built up in a continuous manner through all of the trials. It seems more likely that the learning of the middle syllables begins after the adjoining syllables have been learned" (2, p. 579).

Rather indirect evidence that the serial list is learned piecemeal is found in a study by Obrist (10), who measured the GSR during the course of learning a serial list and found that early in learning the GSR was greatest during the anticipations at the beginning and end of the list and that as the *S* approached the criterion of mastery the GSR became greatest during the anticipations in the middle of the list. From this it does not appear that the *S* was attending to or trying equally hard on every item in the list on every trial, or that he was having greater "difficulty" with the middle items in the series as a result of the development of inhibition or interference during the course of learning. The GSR followed the course of learning and "radiated" out from the first item or anchor point on successive trials.

The assumption concerning the effect of memory span on the skewness of the serial-position curve receives support from an experiment by the investigator to be reported in detail elsewhere. It was found that the serial-position curve of *S*s with a large memory span, as measured independently of the serial learning task, was more skewed or humped to the right than was the curve for *S*s with a small memory span. Also, *S*s who had been trained on a large number of serial learning tasks produced serial-position curves at the end of training that were more skewed than were the curves derived from their first attempts at serial learning.

The theory holds that memory span has less effect the longer the list. Support for this assumption is found in an experiment on memory span by Waugh (13), who showed that "When a series is relatively short . . . a relatively large number of consecutive items are recalled from either end of the series. When a series is relatively long . . . the terminal and initial spans are relatively short. The terminal span seems in this case to be independent of list length . . ." (13, p. 78). These are the very conditions stated by our theory to account for the greater skewness of short serial lists, i.e., lists of fewer than about eight items.

In comparing the serial-position curves of *S*s of superior intelligence (IQ 120-139) with the curves of subnormal *S*s (IQ 40-59), Barnett, Ellis, and Pryer (1) found no over-all difference in the degree of bowing of the serial curves, but did find a difference significant at the one per cent level in the percentage of errors at the second position in a 1-item list. The mentally retarded *S*s had a significantly greater proportion of errors on the second item in the list. The superior *S*s had a significantly greater proportion of errors at the fifth position. Since we know from the subtests of the Stanford-Binet and Wechsler intelligence tests that mentally retarded *S*s generally have a short memory span as compared with intellectually superior *S*s, these differences in the shape of the serial-position curve are exactly what the proposed theory would lead us to expect.

Similar differences in relative errors in the first two or three positions were shown by Malmo and Amsel (7) in comparing the serial-position curves of psychiatric patients having symptoms of severe anxiety with the curves of normal *S*s. Since there is evidence that anxiety adversely affects immediate memory span (11, p. 185; 12), the theory would predict a greater proportion of errors on the first two or three positions for the anxious as compared with the nonanxious *S*s. This is exactly what Malmo and Amsel found.

The theory also states that differences in the degree of flatness or peakedness of the serial-position curve are associated with the degree of agreement among *S*s in the order of learning the items. If the anchor point is somehow obscured or if peculiarities in the composition or presentation of the list increase the effects of idiosyncratic associations, fluctuations in attention, etc., there should be less unanimity among *S*s in their order of learning. Workman (14) obscured the "anchor" points in the list by omitting the first and last items after the first trial, which apparently had the effect of weakening the usual anchor point and of enhancing a more idiosyncratic choice of anchors. The serial-position curve that resulted was quite atypical and nondescriptly irregular. Glanzer and Peters (3) presumably varied the degree of accentuation of the anchor point by varying the duration of the intertrial interval from 0 to 3 to 12.5 seconds, with a presentation rate of 1.5 seconds per item. They found that "As spacing increases, the [serial-position] curve moves from a relatively flat, irregular curve toward the classic bow-shaped curve. The effect, as evaluated by the interaction of spacing with serial position, is significant at the .001 level" (3, p. 8). According to our theory the agreement among *S*s in order of learning should increase with increased spacing, which emphasizes the anchor point, and the increase in agreement among *S*s in order of learning would result in the increasingly bowed curve. Glanzer and Peters determined the degree of agreement in order of learning by means of Kendall's coefficient of concordance; for the three conditions of intertrial spacing (0, 3, and 12.5 seconds) the *W*s were .14, .43, and .47 respectively. Glanzer and Peters concluded that "These coefficients indicate that the individual curves within a group were becoming more alike as spacing increased" (3, p. 9).

E. SUMMARY

A theory was proposed to explain the bow-shaped serial-position curve. The theory is based on five assumptions for which some empirical support is offered: (*a*) the associative connections between the items are learned in a one-trial, all-or-none fashion; (*b*) connections are learned in a certain order around an "anchor" point, which is usually the first item presented; (*c*) the order is determined by the relative ease of forming an associative connection with an already learned item as compared with an item that is unfamiliar or not "anchored" as to serial position; (*d*) the "effort" required by a given *S* to learn the associative connections between items is equally divided among all items in the list; (*e*) the degree of skewness of the serial-position curve is a function of the interaction between the length of the series and the *S*'s immediate memory span.

REFERENCES

1. BARNETT, C. D., ELLIS, N. R., & PRYER, W. Serial position effects in superior and retarded subjects. *Psychol. Rep.*, 1960, **7**, 111-113.
2. BOLLES, R. C. The effect of altering the middle of the list during serial learning. *Amer. J. Psychol.*, 1959, **72**, 577-580.
3. GLANZER, M., & PETERS, S. C. Re-examination of the serial position effect. Technical Report, Research and Development Division, Office of the Surgeon General, Dept. of the Army, Sept., 1960.
4. HAYES, K. J. The backward curve: A method for the study of learning. *Psychol. Rev.*, 1953, **60**, 269-275.
5. HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511-525.
6. JENSEN, A. R., COLLINS, C. C., & VREELAND, R. W. A multiple S-R learning apparatus. *Amer. J. Psychol.* (In press.)
7. MALMO, R. B., & AMSEL, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. Exp. Psychol.*, 1948, **38**, 440-454.
8. MCCRARY, J. W., JR., & HUNTER, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131-134.
9. MCGEOCH, J. A., & IRION, A. L. The Psychology of Human Learning. New York: Longmans, Green, 1952. Pp. 596.

10. Obrist, W. D. Skin resistance and electroencephalographic changes associated with learning. *Summaries of Doctoral Dissertations,* Northwestern University, 1950, **18**, 607-610.
11. Rapaport, D. Diagnostic Psychological Testing. Vol. I. Chicago: Year Book, 1946. Pp. 573.
12. Siegman, A. W. The effect of manifest anxiety on a concept formation task, a non-directed learning task and on timed and untimed intelligence tests. *J. Consult. Psychol.,* 1956, **20**, 176-178.
13. Waugh, N. C. Serial position and the memory-span. *Amer. J. Psychol.,* 1960, **73**, 68-79.
14. Workman, W. G. An experimental investigation of cognitive factors as contrasted with noncognitive factors in rote serial learning. Unpublished Doctoral dissertation, University of Chicago, 1951.

Department of Education
University of California
Berkeley, California

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 143-153.

STIMULUS DISCRIMINABILITY IN VERBAL CONDITIONING*[1]

Department of Psychology, Purdue University

ALBERT R. MARSTON, FREDERICK H. KANFER, AND JOHN F. MCBREARTY

A. INTRODUCTION

Recent studies (Kanfer and McBrearty, 1961, Kanfer and Marston, 1961a, b) have indicated the importance of stimulus variables in verbal conditioning procedures. It has been suggested that tasks such as the one developed by Taffel (1955) consist primarily of discrimination between a presented set of S^Ds and S^Δs rather than of pure operant conditioning. If such were the case, the distinctiveness of discriminative stimuli would be an important variable in determining amount and rate of learning on a given task. Also, as in other discrimination learning, one would predict a decreasing amount of transfer as stimulus discrimination becomes more difficult. Further, when the discrimination of a stimulus dimension is impossible, zero transfer would be predicted.

In an earlier study, stimulus discrepancy was varied using a dimension of degree of hostile connotation (Kanfer and McBrearty, 1961). Hostile verbs were paired with neutral verbs and *S* was reinforced for the selection of the hostile words. The prediction of greater learning with more discriminable stimuli (i.e., intensely hostile paired with neutral verbs) was only partially supported. In a study by Kanfer and Marston (1961a) it was found that with a social reinforcer ("good") *S*s receiving mildly hostile stimuli showed greater learning; whereas with a light as reinforcer the expected superiority of intensely hostile stimuli was evidenced. It was speculated that factors relating to the implication of using intensely hostile words in a social situation confounded the relationship of stimulus discrepancy to learning.

The present study examines further the relationship between stimulus discrepancy and learning using a stimulus dimension which is apparently free of the social implications of hostile words and on which discrepancy can

---

* Received in the Editorial Office on September 13, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This research was supported in part by Research Grant M-3700(C1) to one of the authors (F.H.K.) from the National Institute of Mental Health, U.S. Public Health Service. The authors wish to acknowledge the assistance of Kenneth D. Craig who served as *E*.

be more precisely measured. Nonsense syllables of varying association values were used as stimulus pairs. In terms of the task used, the stimulus dimension of association value was further specified as "similarity to actual English words." Four levels of stimulus discrepancy were presented to four groups by the pairing of: (*a*) high and low association syllables (H-L); (*b*) high and medium (H-M); (*c*) low and medium (L-M); (*d*) and medium and medium (M-M). It was hypothesized that learning and transfer in a verbal conditioning task would decrease with a decrease in stimulus discrepancy.

In studies of verbal conditioning the printed word stimuli are of the same semantic form as the response elicited from *S*. Therefore any manipulation of stimulus variables affects corresponding responses. In the present study, though the stimuli are not actual English words and though *S* was asked to spell rather than pronounce them, variation in the syllables would be expected to influence response probabilities, independent of any reinforcing events. One might predict group differences in responding on an initial, non-reinforced block because of the discrepancies between stimuli selected for conditioning and because of the absolute association value of the syllables. Thus, the frequency of use of syllables which have minimal similarity to English would be expected to be low on the first block, and more so with the greater disparity in the pairs of the H-L group. By the same token high association syllables should have a greater frequency in the H-L than in the H-M group. One would expect approximately equal selection of stimuli in the M-M group and equal selection of the lower association value syllables in the H-M and L-M groups if the absolute association value of the syllables had no effect.

Because of the response bias against the use of lower association syllables, increasing the selection of these syllables by reinforcement would be more difficult. However, the reinforcement of responses with lower initial probability of occurrence allows for a wider range of change during acquisition. Although the absolute association values of the stimuli perhaps affect acquisition as well as initial response probabilities, specific predictions about learning are based primarily upon the relative association values, i.e., the discrepancy between the stimuli.

1. Although the initial frequency of the lower association syllable in H-L is expected to be lowest, this group's block means should be higher and its slope greater than any of the groups with smaller discrepancies between stimulus pairs.

2. Since stimulus discrepancy is approximately equal in the H-M and L-M groups, learning rate and acceleration should not differ; though a lower initial

frequency of the critical association syllable in L-M would tend to lower the level of responding throughout the learning curve.

3. Although a higher initial critical response frequency would tend to elevate the level of its response curve, the M-M group should show lower learning rate than the other groups.

B. METHOD

1. *Apparatus*

*S*s sat at a table in a darkened room facing a 4 × 4 feet masonite board with a 6-inch square window at eye level. Stimulus pairs were typed on 4 × 5 white cards and presented manually by *E*.

2. *Subjects*

Sixty undergraduate *S*s were recruited from elementary psychology classes. Thirty-four were females and 26 males. They were randomly assigned to one of the four experimental groups.

3. *Procedure*

Selection of stimulus pairs. Prior to using nonsense syllables in a verbal conditioning task Glaze's association values (Hilgard, 1951, pp. 541-544) were adjusted to the subject population and task to be used. Fifty pairs of syllables from each of four stimulus discrepancies (H-L, H-M, L-M, M-M) were selected from the Glaze lists. High association value syllables were taken from the 100 per cent section of Glaze's list, Low from 0 or seven per cent, and Medium from 47 or 53 per cent association values. The 200 pairs, randomly arranged on printed data sheets, were presented to 222 *S*s who were instructed to circle the word in each pair which *was more similar in meaning to a real English word.*

The following criteria were used to select from each set of 50 pairs stimuli which best fulfilled the requirements of predetermined levels of stimulus discrepancy. Those pairs for which 90-100 per cent of the 222 *S*s selected the High syllable were assigned for use in the H-L group. The H-M and L-M groups received those pairs for which approximately 75 per cent selected the higher association value syllable. Those pairs for which approximately 50 per cent selected each of the two medium association value syllables were used in the M-M group. The resulting four sets of 20 pairs then had the following mean percentages of preference for the syllable with the higher association value (or in the M-M group for the arbitrarily selected critical member of each pair): H-L = 93.90 per cent, H-M = 74.39 per cent, L-M

= 75.68 per cent, and M-M = 48.95 per cent. The standard deviations ranged from 2.3 to 6.4 per cent.

4. *Verbal Conditioning Task*

Instructions and procedure (similar to those used in an earlier study (Kanfer and Marston, 1961a) were designed to reduce ambiguity in the task. *S*s were told:

> I am going to show you cards with pairs of syllables on them. For each card I want you to choose one of the syllables and spell it. Each time I say "Good" it will mean you selected the correct one.

For each *S*, ten pairs were presented in random order constituting one block of trials. On each subsequent block order of presentation was again randomized. After one block of ten non-reinforced trials, *E* reinforced use of the critical response by saying "Good." In the H-L, H-M, and L-M groups, the lower association value syllables were reinforced. In the M-M group, one member of each pair was selected randomly and reinforced throughout the 50 acquisition trials. The same syllables were reinforced for all *S*s in the group.

After 60 trials (one non-reinforced plus five acquisition blocks) *S* was told he would see another list of pairs and was asked again to choose one syllable of each pair. The remaining ten pairs from the original set of 20 for that *S*'s discrepancy group were then presented. These ten pairs constituted a transfer task and no reinforcement was given.

At the end of the experimental session *E* asked each *S* six questions designed to evaluate his ability to verbalize the contingency between reinforcement and the correct response class. This awareness questionnaire was similar to those used in earlier studies (Kanfer and Marston, 1961a, b). Answers were recorded verbatim and rated for degree of awareness. One author rated all 60 *S*s, the second rated 16 for an estimate of reliability. Reliability of the ratings was .97, a Pearson *r* comparable to that found in an earlier study (Kanfer and Marston, 1961a).

C. RESULTS

1. *Analysis of the Initial Non-reinforced Block*

An analysis of variance on the frequency of the lower association value syllable (or the arbitrarily selected syllable for the M-M group) prior to reinforcement yielded an *F* for treatments of 14.44. As hypothesized, the H-L *S*s spontaneously used very few syllables which have low similarity to English words (mean = 1.00/10). Where discrepancy is smaller in L-M, Low syllables were used more frequently (2.93/10.00). The M-M *S*s selected each syllable with equal frequency (5.00/10.00). As expected on the basis of discrepancy, High syllables were chosen less frequently in the H-M group than they were in the H-L group. However, the frequency of selection of Medium syllables (4.47/10) was somewhat greater than in the standardizing procedure for determining discrepancy.

2. *Over-all Comparisons of Stimulus Discrepancy Level*

The analysis of trends on frequency of critical responses supports the hypothesis relating learning to stimulus discrepancy. The groups differ significantly in terms of over-all performance (F for means over blocks = 4.95), rate of acquisition (F for linear trend = 12.17), and shape of learning curve (F for quadratic trend = 10.51). On each of these measures the group receiving the most widely discrepant stimulus pairs (H-L) has the highest mean value (see Table 2). Further, the group (M-M) receiving stimuli not at all discrepant on the *similarity to English* dimension is consistently lowest on these three measures of learning trend. When tested against zero, all group linear trends are significant. Only H-L has a significant quadratic trend. As was predicted, despite an initial response bias against the use of Low association syllables, the H-L *S*s learn more and learn faster than the groups receiving less discrepant stimuli. Also, the converse is true; the M-M group, despite an initial *advantage* in critical response strength, learns less and with

TABLE 1
SUMMARY OF TREND ANALYSES

	Over-all means		Linear-trend		Quadratic trend		Cubic trend	
Source	MS	F[1]	MS	F	MS	F	MS	F
Treatments	298.67	4.95*	.876	12.17**	.410	10.51**	.066	1.22
Within groups	60.34		.072		.039		.054	

[1] Degrees of freedom for all F tests = 3 and 56.
* $p < .005$
** $p < .001$

TABLE 2
MEANS FOR EACH GROUP FOR THE COMPONENTS OF THE TREND ANALYSIS

	Sum over blocks	Linear coefficient	Quadratic coefficient	Cubic coefficient
H-L	44.67	.811	−.393	+.093
H-M	38.80	.377	−.051	−.043
L-M	33.27	.434	−.060	−.009
M-M	37.33	.248	−.078	−.052

a slower rate. Figure 1 shows the course of learning for each group with the response frequency on each block adjusted for the group's mean response frequency on the initial, non-reinforced block.

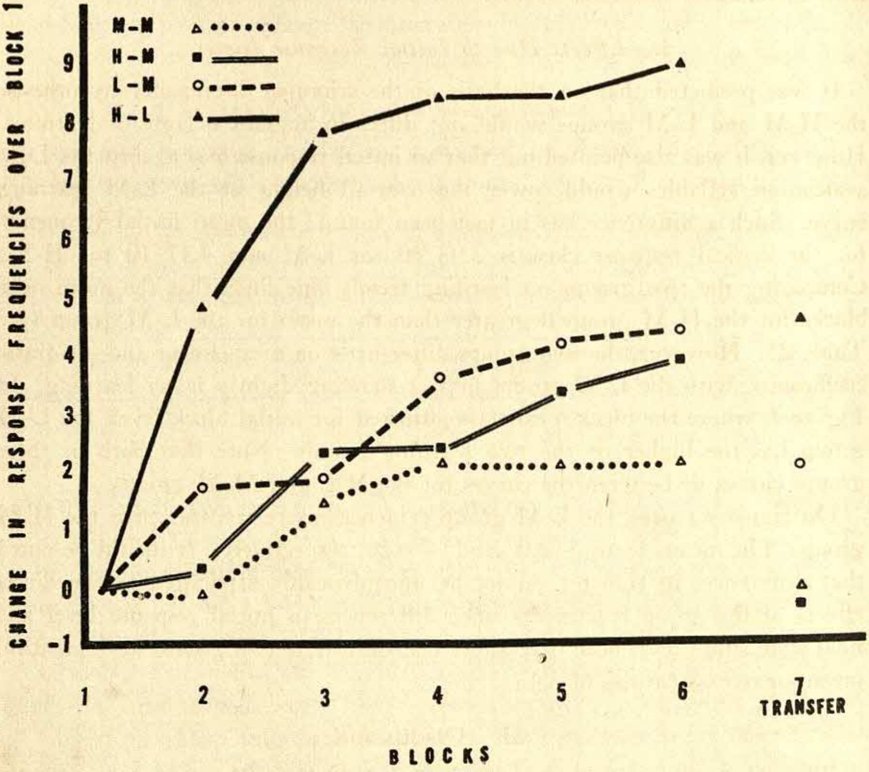


FIGURE 1
Improvement over Pre-Acquisition for all Groups

Findings on analyses of transfer scores and awareness ratings also support the discrepancy hypothesis. Amount of transfer was defined as the difference in frequency of critical responses between the initial preacquisition block and the block on a second list. The F for treatments on an analysis of variance on these difference scores was 9.00 ($df = 3, 56, p < .001$). The means increased in the following order: H-M (— 0.20), M-M (— .07), L-M(2.07), H-L (4.13). As with learning trends, the H-L group shows significantly greater transfer than the other groups.

The groups differed significantly on awareness ratings ($F = 10.38$). Again H-L shows a greater mean score than any of the other groups and

M-M the lowest mean score. The means are: H-L (5.40), H-M and L-M (3.13), M-M (2.67). The awareness ratings, based on ability to verbalize about the correct response class, correlate significantly (Pearson $r = .50$) with the transfer difference scores.

3. *Effects Due to Initial Response Biases*

It was predicted that on the basis of the stimulus discrepancy hypothesis, the H-M and L-M groups would not differ in amount or rate of learning. However, it was also pointed out that an initial response bias against the Low association syllables would lower the over-all height of the L-M learning curve. Such a difference has in fact been found; the mean initial frequency for the critical response class is 2.93/10 for L-M and 4.47/10 for H-M. Comparing the two groups on learning trends, one finds that the mean over blocks for the H-M group is greater than the mean for the L-M group (see Table 2). However, the two groups differ little on mean linear and quadratic coefficients, with the L-M groups in fact showing slightly faster learning. In Figure 1, where the block means are adjusted for initial block level, the L-M group has the higher of the two learning curves. Note that both of these groups curves lie between the curves for the H-L and M-M groups.

On transfer scores, the L-M group evidences more transfer than the H-M group. The means were + 2.07 and — 0.20, respectively. It should be noted that differences in transfer cannot be unequivocably attributed to the direct effects of the group treatments since differences in initial response level and final acquisition level both may affect transfer. The two groups have identical mean awareness ratings of 3.13.

D. DISCUSSION

The results support the hypothesis that learning and transfer in a verbal conditioning task decrease with a decrease in stimulus discrepancy. In all analyses, the H-L group, i.e., those *S*s receiving the most widely discrepant stimuli, evidence higher performance. Conversely, the M-M group, those *S*s who received stimulus pairs not discriminable on the similarity to English dimension, showed poorest performance on the measures of learning and awareness. These group differences are particularly clear evidence in favor of our hypothesis in light of the H-L group's bias against the use of the critical responses on the block of trials preceding acquisition.

As mentioned previously, in earlier studies (Kanfer and McBrearty, 1961; Kanfer and Marston, 1961a, b) the relationship between stimulus discrepancy and learning was only partially supported using hostile connotation as the critical dimension. Our earlier assertion that the content of the response classes influenced learning and our speculation that a dimension with fewer social implications would provide a clearer demonstration of the discrepancy hypothesis appear to have been confirmed in the present study.

In the intermediate range of discrepancy, there is less consistency in the present findings. Although the H-M and L-M groups evidence learning curves with similar slopes and acceleration, the height of the H-M curve is somewhat greater over all blocks. It appears that one cannot deal with stimulus discrepancy simply in terms of distances along a single dimension. Although the H-M and L-M groups received stimuli equally discrepant in magnitude of association value, the H-M *S*s used the critical, lower association syllable more frequently on the initial, non-reinforced block and maintained a greater over-all mean throughout acquisition. Response bias affecting particularly the emission of Low association syllables must be taken into account whenever, as in the present task, variation of stimuli produces concomitant variation in an isomorphic response. This bias may be attributable to the absolute strength of the critical response rather than to the relational aspects of the stimulus pairs. The results indicate that where groups vary widely in stimulus discrepancies, as in the H-L and M-M comparisons, differences in learning correspond to the predictions based upon the stimulus discrepancy hypothesis. Within a narrower range of discrepancy, response bias apparently weighs more heavily.

It was argued earlier (Kanfer and Marston, 1961a,b) that the Taffel type of verbal conditioning procedure mainly involves discrimination learning rather than free operant conditioning. The present findings emphasize the effect of the presented S^Ds and S^Δs upon amount of ambiguity in this type of conditioning situation. In the group where discrimination on a stimulus dimension was not possible (M-M) minimal learning was evidenced and positive transfer did not occur; whereas, the availability of such a dimension in the H-L group allowed for considerable learning and positive transfer. *S*s' verbalizations also reveal the importance of the stimulus variables as a source of ambiguity. In the M-M group phrases such as "they didn't seem to make sense," "I just remembered the word and said it again," and "it was just kind of arbitrary" revealed the *S*s' searching for relevant dimensions. The fact that the group showed a significant increase in the frequency of critical responses over blocks indicates that *S*s were able to learn specific responses in lieu of utilizing a dimension.

Adopting the position taken in earlier studies (Kanfer and Marston, 1961a, b), *S*s' verbalizations about the response-reinforcement contingencies were treated as a dependent variable and rated as to degree of "awareness."

One problem associated with use of verbal reports is the dependence of "awareness" upon the nature of the questioning procedure. Four questions were used in the present study; a general one, "what is the purpose of the experiment?", followed by three more specific questions concerning the selection of responses. *S*s in the extreme groups (H-L and M-M) were rated twice; once on just the reply to question one and a second time using the verbalizations on the more specific queries. The H-L group showed a significant increase in awareness ratings when the specific questions were evaluated ($t = 2.43$, $p < .05$). The mean on the general question was 3.96; the mean on the specific questions 5.53. It was also noted that though *S*s in all groups were given the information necessary to achieve a two rating, i.e., they were told "Good means you have selected the correct syllable," eight of them received ratings of one, even when the more specific questions were asked. In the M-M group awareness ratings increased from a mean of 1.67 on the general question to a mean of 2.73 on the specific questions ($t = 3.70$, $p < .01$). Further, six out of 15 *S*s in the M-M group received ratings of four on the specific questions. Although none of the *S*s in the group received ratings greater than two on the general question, when asked the more specific queries 53 per cent of them verbalized awareness of the dimension which was correct for all but the M-M *S*s. These findings indicate that not only do the questions themselves influence awareness responses, as noted by Levin (1961), but that verbalizations about a correct dimension can occur despite the fact that the range of stimuli available during conditioning seems to preclude such "awareness" in the M-M group.

These results of analyzing awareness ratings can perhaps best be handled by treating "awareness" as a group of verbal responses which are available in the *S*'s repertoire and which vary in probability depending upon various stimulus conditions. Thus, despite the instructions, awareness responses rated one had a low but not zero probability of occurrence. Similarly, in the M-M condition responses rated greater than two had a low but not zero probability of occurrence and increased in probability when more specific questions changed the stimulus conditions.

Levin's interpretation (1961) that learning does not occur without awareness raises the spectre of the alternate view that awareness does not occur without learning. It appears that the present analysis in terms of response probabilities obviates the necessity for solving this dilemna and permits more adequate examination of the empirical relationships of awareness responses to conditioning.

The conceptualization of awareness as a response which has varying probabilities of occurrence depending upon stimulus conditions is consistent with the problem solving approach to verbal conditioning set forth in recent articles (Adams, 1959; Kanfer and McBrearty, 1961; Kanfer and Marston, 1961a, b). Verbal conditioning is conceptualized as several concurrent interdependent learning tasks relating to the discovery of the nature of the experiment. Various hypotheses are tested by *S* in his efforts to solve the problem of the nature of the task confronting him. The present study discusses awareness as the probability of occurrence of correct hypotheses as they are expressed in verbal responses.

Previous studies have discussed the analogy between verbal conditioning and psychotherapy. The results of the present study indicate that if such an analogy is to be made, the importance of discrimination learning in the verbal conditioning task and of the effect of stimulus variables must be considered.

E. SUMMARY

The present study investigated the effect of the distinctiveness of discriminative stimuli upon the amount and rate of learning in verbal conditioning. In a pre-experiment task 222 *S*s rated pairs of nonsense syllables in terms of *similarity to English*. Four lists of stimulus pairs, representing four levels of stimulus discrepancy (as defined by Glaze association values and the results of the pre-experiment task) were selected: High-Low, High-Medium, Low-Medium and Medium-Medium. Each of four groups received a set of 10 pairs, taken from one of the discrepancy levels, in a modified verbal conditioning task. *S*s were asked to select one member of each pair and spell it. After the 10 pairs were presented in a pre-acquisition block, they were repeated in random order for five subsequent blocks, with the lower association syllable reinforced by *E*'s saying "Good" (in Medium-Medium an arbitrarily selected critical response was reinforced in each pair). After the 60 trials, a second set of 10 pairs, taken from the same discrepancy level, was presented on a transfer task with no reinforcement. A questionnaire designed to test *S*'s ability to verbalize the contingency between *E*'s reinforcement and the use of responses from the correct class was administered after the transfer task. The results supported the hypothesis that increased stimulus discrepancy yields a higher level and faster rate of learning. In a trend analysis on the frequency of critical responses per block, the High-Low group showed the highest overall mean, linear trend, and quadratic trend; the Medium-Medium group was lowest on all of these measures. All *F*s were significant at the .005 level. High-Low *S*s' were also highest on transfer scores and ratings of awareness

of the response-reinforcement contingency. The groups receiving intermediate levels of discrepancy (High-Medium and Low-Medium) yielded learning curves lying between those for the High-Low and Medium-Medium groups. The results were discussed in terms of the role of stimulus discrepancy as a source of ambiguity in verbal conditioning.

REFERENCES

1. ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383-405.
2. HILGARD, E. R. Methods and procedures in the study of learning. *In* S. S. Stevens (*Ed.*), *Handbook of Experimental Psychology*. Wiley & Sons, Inc., New York, 1951.
3. KANFER, F. H., & MARSTON, A. R. Verbal conditioning, ambiguity and psychotherapy. *Psychol. Rep.*, 1961, **9**, 461-475. (a)
4. ———. The effects of task-relevant information on verbal conditioning. *J. Psychol.*, 1962, **53**, 29-36. (b)
5 KANFER, F. H., & MCBREARTY, J. F. Verbal Conditioning: Discrimination and awareness. *J. Psychol.*, 1961, **52**, 115-124.
6. LEVIN, S. M. The effects of awareness on verbal conditioning. *J. Exp. Psychol.*, 1961, **61**, 67-75.
7. MATARAZZO, J. D., SASLOW, G., & PAREIS, E. N. Verbal conditioning of two response classes: Some methodological considerations. *J. Abnorm. Soc. Psychol.*, 1960, **61**, 190-206.
8. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. Abnorm. Soc. Psychol.*, 1955, **51**, 496-501.

Department of Psychology
Purdue University
Lafayette, Indiana

Published as a separate and in The Journal of Psychology, 1941, 11, 155-166.

INCIDENTAL LEARNING: A FUNCTION OF ASSOCIATIVE STRENGTH AND DISTANCE BETWEEN S-R PAIRS*

Department of Psychology, Washington Square College, New York University

M. S. [illegible] AND M. E. [illegible]

[illegible]

[illegible]

[illegible]

---

* Received in the Editorial Office on [illegible], 1941, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[illegible] New York University [illegible] Fund.

Published as a separate and in *The Journal of Psychology,* 1961, **53**, 155-160.

INCIDENTAL LEARNING: A FUNCTION OF ASSOCIATIVE STRENGTH AND DISTANCE BETWEEN S-R PAIRS*[1]

University of Southern California and New York University

M. S. Mayzner and M. E. Tresselt

A. Introduction

It has been shown for intentional learning (1, 2) that the immediate free recall of a list of words is a function, in part, of the associative strength between word-pairs in the list. Specifically, *S*s were read a list of words once and then were asked for their immediate free recall. The word lists were comprised of stimulus-response pairs selected from the Kent-Rosanoff word list and were constructed to vary systematically in the mean associative strength between *S-R* pairs, as measured by frequency of occurrence in word association norms. It was found, in general, that Kent-Rosanoff pairs tended to be recalled together in an *S-R* or *R-S* sequence (associative clustering) and that the amount of such associative clustering was a function of the mean associative strength of the *S-R* pairs in each list.

The present study is concerned with examining the effects of associative strength and distance between *S-R* pairs in an incidental learning task. It is predicted that word lists comprised of *S-R* pairs with high associative strength should yield greater recall scores and show more associative clustering than word lists comprised of *S-R* pairs with low associative strength. Also, greater recall scores and more associative clustering should occur in lists in which the distance between the *S* and *R* of an *S-R* pair is zero, i.e., no intervening items occur between the *S* and *R* of an *S-R* pair, in contrast to lists in which the distance is greater than zero, i.e., some specific number of intervening items occur between the *S* and *R* of an *S-R* pair.

B. Method

1. *Experiment I*

*a. Subjects. S*s were 75 students from introductory psychology classes at New York University.

---

* Received in the Editorial Office on September 20, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was supported in part by a grant made to M. E. Tresselt by the New York University Arts and Science Research Fund.

b. Stimulus material and procedure. Three separate word lists were constructed from the stimulus and response items presented in Table 1. Each list was comprised of the 15 stimulus words and 15 response words to form a high, medium, and low associative strength list, i.e., the words in the high list were table-chair, dark-light, etc.; in the medium list, table-food, dark-night, etc. The 30 words in each list were typed with double-spacing between words in a single column on a white 8½ × 11-inch sheet of paper. Since each of the three lists were presented to a different group of 25 *S*s, 25 different random orders of stimulus and response words were constructed as follows. Each stimulus word and each high response word were typed on a separate card. This pack of 30 cards was then shuffled and the resulting ordering constituted the first random array for the high list condition, with random orders for the medium and low list conditions being derived from the high list condition by merely substituting a medium or low response word where each high response word had appeared. This procedure was repeated 25 times to yield 25 different random orders to be applied across the high, medium, and low word lists, yielding a total of 75 separate 30 word lists, which were randomly distributed to the 75 *S*s employed in Experiment I.

TABLE 1
KENT-ROSANOFF PAIRS USED IN EXPERIMENTS I AND II

Stimulus	Response words					
words	High	f*	Medium	f	Low	f
1. Table	chair	840	food	41	book	1
2. Dark	light	829	night	55	red	1
3. Soft	hard	445	loud	22	heart	1
4. Black	white	751	cat	26	big	1
5. Rough	smooth	439	road	44	wind	1
6. Spider	web	454	leg	50	corner	1
7. Needle	thread	464	sharp	55	record	1
8. Girl	boy	704	woman	49	baby	1
9. Sour	sweet	568	bitter	60	neighbor	1
10. Stem	flower	402	root	47	ground	1
11. Hammer	nail	537	head	50	board	1
12. Bed	sleep	584	sheet	36	floor	1
13. Tobacco	smoke	515	pipe	49	field	1
14. Scissors	cut	671	paper	43	hair	1
15. Salt	pepper	430	taste	50	drink	1
Mean		575		45		1

* Frequency norms from Russell and Jenkins (3).

All testing was conducted during regular class periods, at which time, employing a technique developed in an earlier study (4), *E* informed the class he was interested in obtaining samples of handwriting, that lists of words would be handed out, and that each *S* was to copy the list using his normal handwriting. At the completion of this task all lists and handwriting samples were collected and *S*s were then asked for their immediate free recall of the list they had just copied, with 10 minutes being allowed for recall.

2. *Experiment II*

a. Subjects. *S*s were 150 students from introductory psychology classes at New York University.

b. Stimulus material and procedure. Six separate word lists were constructed from the stimulus and response items presented in Table 1, with each list being presented to a different group of 25 *S*s. Three lists used the 15 stimulus words and the 15 high response words and another three lists used the 15 stimulus words and the 15 low response words. In each set of three lists (high associative strength and low associative strength) one list had zero distance between the *S* and *R* of an *S-R* pair, the second list had a distance of three, and the third list had a distance of five, i.e., zero, three, or five intervening items occurred between the *S* and *R* of an *S-R* pair. In the lists with zero distance, for both the high and low associative conditions, 25 different random orders of the 15 *S-R* pairs were constructed, analogous to the randomization procedure used in Experiment I. Similar randomization procedures also were employed for the high and low associative conditions with distances of three and five items occurring between the *S* and *R* of an *S-R* pair. This resulted in a total of 150 separate 30 word lists which were randomly distributed to the 150 *S*s employed in Experiment II. All testing was done during regular class periods and the instructions and procedure were identical to those used in Experiment I.

C. RESULTS AND DISCUSSION

1. *Experiment I*

The results of Experiment I are presented in Table 2, which shows the mean number of total words, i.e., both *S* and *R* words, recalled and the mean number of associative clusters, i.e., both forward and reverse pairs, for the high, medium, and low associative strength conditions. The mean number of words recalled shows a systematic drop from 13.28 to 9.60 to 9.00, as associative strength decreases, and a simple analysis of variance yields an F of 7.47, with a corresponding P of less than .005, indicating that the effect of associative strength on number of words recalled is highly significant. The mean number of clusters also shows a systematic drop from 1.80 to

.48 to .36, as associative strength decreases, and the application of the Kruskal-Wallis one way analysis of variance test yields an H of 19, with a corresponding P of less than .001, indicating the highly significant effect of associative strength on clustering.

TABLE 2
TOTAL WORDS RECALLED AND NUMBER OF ASSOCIATIVE CLUSTERS FOR HIGH, MEDIUM, AND LOW ASSOCIATIVE STRENGTH LISTS

	High	Medium	Low
Total Words			
Mean	13.28	9.60	9.00
SD	4.75	4.34	3.55
Number of Clusters			
Mean	1.80	.48	.26
SD	1.26	.57	.69

2. *Experiment II*

The results of Experiment II are presented in Table 3, which shows the mean number of total words recalled and the mean number of associative clusters for the high and low associative strength conditions, with *S-R* distances of 0, 3, and 5 items. First, with regard to the mean number of words recalled, a 2 × 3 factorial analysis of variance yields *F*s of 14.66, 2.05, and 6.72, for the main effects due to associative strength, *S-R* distance, and the interaction term, respectively, with corresponding *P*s of less than .001, between .10 and .20, and less than .005, respectively. Thus, the effect of associative strength is highly significant, while the effect of *S-R* distance is not significant. However, since the interaction term is highly significant, simple analyses of variance were performed separately, for the high and low associative strength conditions, yielding *P*s of less than .005 and greater than .20 for the high and low conditions, respectively. From this it may be con-

TABLE 3
TOTAL WORDS RECALLED AND NUMBER OF ASSOCIATIVE CLUSTERS FOR HIGH AND LOW ASSOCIATIVE STRENGTH LISTS

Associative strength		S-R distance 0		3		5	
		Totals	Clusters	Totals	Clusters	Totals	Clusters
High	*M*	14.64	5.84	10.80	1.60	9.72	1.20
	SD	4.84	2.89	4.92	1.62	4.64	1.10
Low	*M*	7.88	.72	9.56	.68	9.08	.16
	SD	3.86	1.69	4.56	.97	4.12	.37

cluded that, while *S-R* distance has no effect when associative strength is low, it has a highly significant effect when associative strength is high, on the total number of words recalled.

With respect to the mean number of clusters, an application of the Kruskal-Wallis one-way analysis of variance test yields *H*s of 33 and 4, with associated *P*s of less than .001 and between .10 and .20, for the high and low associative strength conditions, respectively; thus indicating that *S-R* distance has a highly significant effect on associative clustering when associative strength is high, but no effect when associative strength is low, and is consistent with the effect of *S-R* distance on the total number of words recalled. A comparison of the mean number of clusters between high and low associative strength conditions for 0, 3, and 5 *S-R* distance values all yield highly signficant *P* values and is consistent with the results for the total number of words recalled.

The results of Experiments I and II extend the generality of the effects of associative strength in intentional learning (1, 2) to incidental learning tasks. The findings of Experiment II also indicate that the number of intervening items or words between the *S* and *R* of a particular *S-R* pair significantly affects the total number of words recalled and the number of associative clusters produced, if associative strength is high, but not if it is low, and this interaction of associative strength and *S-R* distance warrants further examination in intentional learning tasks.

D. SUMMARY

The present study examined the effects of two variables, associative strength and distance between *S-R* pairs, in an incidental learning task. The results showed that word lists comprised of *S-R* pairs with high associative strength yield greater recall scores and more associative clustering than word lists comprised of *S-R* pairs with low associative strength. However, associative strength and *S-R* distance were found to significantly interact, producing greater recall scores and more associative clustering when *S-R* distance is small rather than large, for high associative strength lists, but producing no difference in low associative strength lists.

REFERENCES

1. Jenkins, J. J., & Russell, W. A. Associative clustering during recall. *J. Abnorm. Soc. Psychol.*, 1952, **47**, 818-821.
2. Jenkins, J. J., Mink, W. D., & Russell, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127-136.

3. RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Tech. Rep. No. 11,* 1954, under Contract N8-ONR-66216 between the Office of Naval Research and the University of Minnesota.

4. TRESSELT, M. E., & MAYZNER, M. S. A study of incidental learning. *J. Psychol.,* 1960, **50,** 339-347.

Department of Psychology
University of Southern California
Los Angeles 7, California

Published as a separate and in The Journal of Psychology, 1962, 53, 161-181.

STUDYING STUDENTS AS STUDENTS: A CASE EXAMPLE*

Division of Student Mental Hygiene, Department of University Health, Yale University

---

JAMES S. DAVIE AND RALPH M. RUST

A. INTRODUCTION

Although many college students, especially those enrolled in courses in psychology or sociology, have the experience of being subjects in studies conducted by faculty members, few faculty members are interested in studying students as students. Students selected as subjects have generally been viewed as convenient means for testing some theoretical proposition or a proposition which usually has little or nothing to do with the psychological status of the individual as a young adult student in a specific educational environment.

Since World War II, and especially in the past decade, there has been a good deal more of interest in understanding students as students. In addition to the earlier work at Yale ([illegible]), major studies have been undertaken of such institutions as Bennington ([illegible]), Cornell ([illegible]), Vassar ([illegible]), Dartmouth ([illegible]), Harvard, Antioch, Reed, San Francisco State, and Swarthmore. In addition, many other studies limited to selected aspects of the college experience are underway in other colleges. Much of the impetus for this upsurge in interest has come from the Social Science Research Council's Committee on Personality Development in Youth ([illegible]), from foundations such as Carnegie, Hazen and the Danforth, and more recently from research centers such as the Center for the Study of Higher Education at Berkeley as well as from federal agencies such as the National Institute of Mental Health and the Office of Education.

As staff members of a student mental hygiene clinic, who had long been interested and involved in studying students as students, the authors are encouraged to note the gradual development of this important area of study which is legitimate, respectable, and important. In the present paper, we propose to report on a study of college students in which we are currently involved. Our report will be a descriptive one, concentrating on the scope, content, methodology, and rationale of the study but leaving "findings" for future papers. Because of the size of the study, we prefer to use the present paper as a framework for subsequent papers which will discuss in detail

---

* Received in the Editorial Office on September 27, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 161-182.

STUDYING STUDENTS AS STUDENTS: A CASE EXAMPLE*

Division of Student Mental Hygiene, Department of Health, Yale University

JAMES S. DAVIE AND RALPH M. RUST

A. INTRODUCTION

Although many college students, especially those enrolled in courses in psychology or sociology, have the experience of being subjects in studies conducted by faculty members, few faculty members are interested in studying students as students. Students selected as subjects have generally been viewed as convenient bodies for testing some theoretical proposition—a proposition which usually has little or nothing to do with the psychosocial status of the individual as a young adult studying in a specific educational environment.

Since World War II, and especially over the past decade, there has been a gradual increase in interest in understanding students as students. In addition to the authors' work at Yale (5-7, 10-14, 16), major studies have been undertaken at such institutions as Bennington (9), Cornell (8), Vassar (15), Princeton (4), Amherst (2), Harvard, Antioch, Reed, San Francisco State, and Swarthmore. In addition, many other smaller studies of selected aspects of the college experience are underway in other colleges. Much of the impetus for this upsurge in interest has come from the Social Science Research Council's Committee on Personality Development in Youth (1), from foundations such as Carnegie, Grant and Old Dominion, and more recently from research centers such as the Center for the Study of Higher Education at Berkeley as well as from federal agencies such as the National Institute of Mental Health and the Office of Education.

As staff members of a student mental hygiene clinic that has long been interested and involved in studying students as students, the authors are encouraged to note the gradual recognition of this topic as an area of study which is legitimate, respectable and important. In the present paper we propose to report on a study of college students in which we are currently involved. Our report will be a descriptive one, concentrating on the scope, content, methodology, and rationale of the study but leaving "findings" for future papers. Because of the size of the study, we prefer to use the present paper as a framework for subsequent papers which will discuss in detail

---

* Received in the Editorial Office on September 20, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

separate aspects of the study and will present relevant findings. This procedure should eliminate the necessity of repeating in each subsequent paper many of the details of purpose, design, and methodology.

Because the present study is not an isolated venture into the area of research on college students but rather represents one phase of a collaborative research effort over the past decade, we believe that our description of the study and, particularly, our exposition of its details will be of value to those now contemplating such studies for the first time. Judging from our correspondence with faculty at other institutions over recent years, the experience we have accumulated in studying our own students has been relevant and useful for others studying their students. We have also come to recognize that for various reasons our approach (and biases) are somewhat different from those of others. We therefore believe that a systematic description of our views and procedures would serve to illustrate one way of viewing the field. While we do not expect anyone to adopt our approach *in toto,* we do feel that it contains some elements which some researchers will find worth borrowing and other elements which they will be stimulated to improve upon. (We also expect that there will be some whom we stimulate to try something completely different.)

Succinctly stated, our study is a multi-purpose, multi-dimensional study of the college experience based on the perceptions of a representative cross-section of undergraduates in a given college at a given point of time. It is part of an extensive research program maintained by the Division of Mental Hygiene of the Department of Health at Yale University. While the division serves the entire student community, both graduate and undergraduate, the present study is restricted to the undergraduate level. The latter may be best described as a private, residential, liberal arts college for men, of approximately 4,000 students.

B. THEORETICAL ORIENTATION

We believe it both necessary and desirable to specify our theoretical interests at the outset. These may then serve as a frame of reference for description of the questionnaire content and of our methods and analytical procedures.

1. *General Orientation*

Our most general interest is in young adulthood, roughly the age group of the college student. We feel that this segment of the life cycle has distinctive aspects and deserves the same careful attention that social scientists have paid to other segments of the life cycle such as early childhood, adolescence, and old age. Since we are staff members of a psychiatric clinic in a college health service we are particularly interested in the young adult who is in college and in the question of how the college experience affects and reflects the personality of the student. We try to understand our students, whether patients, pupils, or research subjects, in terms of their individual biography, their position in the life cycle, and their membership in a specific educational community and in groups within that community.

Fusing these considerations, we have a tentative working model of the college years as a transitional period during which the student moves from a state of childlike dependence to one of responsible adulthood. In terms of dynamic ego psychology it is a period during which the ego is essentially concerned with the task of mastering external reality and of bringing the individual into some comfortable and meaningful relationship with the world in which he must live. This is in contrast to earlier adolescence where the main task was that of mastery of internal impulses (16). In more sociological terms, it is a period during which the individual is breaking away from his family of orientation and moving toward that imminent point in time when he is expected to function as an emotionally independent, self-supporting, productive and responsible member of society with a family of his own.

During the college years there is an increasing amount of interaction with the peer group which may function as a source of new and a modifier of old experience. The peer group also serves as a source of emotional support for the individual as he lessens his ties of emotional dependency on the family, redefines his relationships to them, and experiments with various roles in an attempt to synthesize from elements in his total range of experience a life role which is tailored to his own needs and to external circumstances as he perceives them (5). Questions of vocational choice (for women, career vs. marriage vs. both) and of heterosexual relations around which so much of adult life revolves are of vital import during this period.

The young adult in college also occupies the status of student and is confronted with the additional problems of adjustment to the college environment. This, broadly conceptualized, consists of two major cultures: the official culture of the faculty and administration and the peer culture of the student body. Each is a more or less organized entity exhibiting shared patterns of belief and behavior. As does the larger society each transcends the individual and imposes requirements upon him to which he must make some kind of response (4). The kind of response he makes to the college experience, the meaning of the college experience to him, and the effect it has upon him will be influenced considerably by what he brings to the college situation. Many of the student's responses will be common to those of others due to selective and restrictive admissions policies and to common experiences he has had in the larger society. Others will be highly idiosyncratic by virtue of unique endowments and/or highly personal experiences.

Whatever the response of a given individual, the college years represent, for most students, a period of time during which the personality may be sufficiently pliable to permit considerable change to occur. They are a period during which many students are away from home for the first time, thus having the first opportunity and first challenge to function relatively independently as well as to seek new sources of social support. During this time there is continued exposure to formal and informal educational influences which on the whole reflect the orientations and values of larger segments of society. During the college years many students begin to concentrate seriously on choosing and preparing themselves for many life roles, which implies the attempt at and consolidation of a personal and social identity. In brief, the life course and more or less final form of the personality may be vitally affected by events of the college years.

2. *Major Focal Points*

Within the framework of the general orientation sketched above, we were specifically interested in five topics: satisfaction with the college experience, academic achievement in relation to ability, student problems, student behavior, and changes occurring during the college years. The first two topics have been subjects of extensive prior investigation by the authors (5, 10).

Briefly, satisfaction with the college experience was viewed as a resultant of the interaction process between student and environment. It was conceptualized as a feeling tone which reflects the success with which the individual handles *both* internal needs and external pressures. The research task was threefold: to measure satisfaction; to demonstrate its relationship to measures of individual success both in utilizing the college environment for his own purposes and in adjusting to the demands made of him by the environment; and to determine what factors other than those hypothesized further differentiated the satisfied from the dissatisfied student.

The interest in academic achievement was focused on factors related to changes in academic achievement (relative to ability) occurring from secondary school to college. Previous work had hypothesized that such changes in achievement are associated with changes in behavior favorable to the acquisition of good grades (e.g., amount of time spent in studying); that changes in behavior are a function of the student's response to the greater freedom from supervision that college provides; that the student's response to this greater freedom is a function of the degree of internalization of parental values conducive to academic achievement; and that the degree of internalization of parental values is a function of the nature of the parent-child relationship. Accordingly, the research task was to measure academic achievement and demonstrate its relationship to such variables as study habits, time use, parental expectations, and the student's own values.

The third focal point, student problems, represented a special interest of the authors in the stresses and strains of the college experience (11, 12). On the assumption that much about the normal can be learned from a study of the abnormal or of those individuals having difficulty in a given situation, we wanted to pay particular attention to the kinds of problems experienced by different kinds of students. Students who consulted the mental hygiene service were obviously an excellent source of data on this topic (6). However, while much can be learned from students who take their problems to such official agencies of the college, analysis of data from these sources inevitably yields an incomplete picture. There always remains the question of those students with problems who have not come to the attention of agency officials. Accordingly, we wanted to examine the question of problems in the general student population as well as in that part of the population that utilized the clinic.

The fourth focal point, student behavior, was more a methodological than a theoretical concern. Impressed in our previous studies by the failure of traditional attitude-type items and scales to differentiate various study groups and the comparative success of concrete behavioral items in doing so, we undertook to measure with a Reported Behavior Inventory as many aspects of the college experience as we could meaningfully reduce to such terms (13, 14). It was our hope that the behavioral items would prove useful both in the specific explorations of satisfaction and academic achievement as well as in the documentation of our general model of the college years.

The final focal point, changes occurring in time, is self-explanatory. We were interested in the question of differences in the degree of satisfaction and relative academic achievement and their correlates at different points in the college years. Similarly, we hoped to describe general changes occurring in other aspects of the college experience not encompassed by the topics of satisfaction and achievement. We were particularly interested to see if we could do so in behavioral terms, as we were somewhat dissatisfied with other studies of personality change during college which demonstrate changes in response to attitude items but without much, if any, evidence of such changes being reflected on the level of actual behavior. The study was not a longitudinal one in the sense that it studied changes in the same individuals over time. Rather it was viewed as a preliminary step to conducting such a study. By examining different variables and their inter-relationships at the end of each of the four college years, we hoped to approximate a longitudinal study and to alert ourselves to those variables which might prove most fruitful to study in the same individuals in a subsequent study. As yet we have not undertaken a true longitudinal study, but we do have longitudinal data on some subjects from the present study. The same questionnaire was readministered to those seniors who had responded as freshmen and who were still in college three years later.

C. Methods and Techniques

To achieve the theoretical objectives described above, a lengthy (twenty-six page) questionnaire was constructed and administered to a sample of undergraduates in the spring of 1957.

1. *The Questionnaire*

We used a questionnaire as our basic method for data collection because it provided a relatively inexpensive and quick way of collecting the great amount of data we needed. While we do not recommend a questionnaire as an exploratory device (its efficient use presupposes that the researcher has at least a rudimentary grasp of the topic under examination and knows what questions to ask and how to ask them), we felt our use of a questionnaire was justified on the basis of extensive past experience with questionnaire research on our students. From the results of our previous questionnaire studies, from our clinical experience, and from the results of other studies using different data collection techniques we had accumulated sufficient knowledge of the subject matter to feel that for our purposes the questionnaire was the most appropriate device to use. In addition, previous experience had indicated that our students would participate in such questionnaire studies if properly approached. Generally speaking, they would respond if convinced of the importance of the study and of their participation in it, if interested in the subject matter of the questionnaire, if not antagonized by impertinent or ambiguous wording of the questions, and if convinced of the researchers' integrity. In fact, the more the questionnaire appeared to be focused on the realities of their campus and their personal life situation, the more likely were they to respond. For this reason, we have generally preferred to construct questionnaires tailored to our own situation rather than to rely on standardized instruments. Our students often regard such standardized instruments as artificial, distortive, and, where designed for administration to the general population, often insulting to the intelligence of the college student.

While tailoring of research instruments to a specific local situation creates certain problems of comparability of results, it is our belief that the first task of the researcher is to develop a thorough knowledge of *his* students and *his* institution. This belief is partially a reflection of our experience as staff members of a *service* department which daily must deal with the realities of our own campus. We have found that the more specific and locally relevant are the data, the greater is the usefulness of such data from a service standpoint. It is also a reflection of our more general conviction that the understanding of similarities and differences pertaining between institutions presupposes intensive study and knowledge of *each* individual institution.

Despite the individualized design of our research instrument we focus on many of the common denominators of college life whatever the college and, as implied earlier, have contributed many items to others studying their own colleges. Similarly, we have been quick to borrow items from others where those items promised to be useful and relevant for our own work.

Our questionnaire then was not a standardized instrument but did contain many standardized questions. These were standardized internally in the sense of having been asked of many successive samples of our own students and externally in the sense of having been asked by others of students in their colleges.

The fact that our questionnaire contained so *many* questions, whether standardized or not, is a reflection not so much of our unbridled curiosity but rather of our desire to study *simultaneously* as many different aspects of the college experience as possible and in the *same* individuals. Most studies of college students as students involve the inter-relationships of a few variables at a single point in time. The literature on academic achievement is an excellent example of this. Such studies generally have a few specific hypotheses and a narrow focus of interest. Data which may be of relevance to the topic at hand but are not hypothesized in advance are generally overlooked. Even where a single study of one limited aspect of college is part of a program or larger study involving other small studies of other limited aspects, the fact remains that such *separate* studies seldom use the same subjects. By contrast, our intent was to examine the interrelationships of as many aspects of college as possible *within the same subjects at the same point in time.*

The difficulty with such a desire is, of course, that to execute it properly requires a great investment of time and energy. It also tends to make the investigators suffer from acute intellectual indigestion as they face such a quantity of data. The indigestion is particularly acute when one uncovers significant results for which no *a priori* expectation or hypothesis exists. However, our alkalizer for such discomfort is generated by the data themselves. They quickly induce an increased awareness of the complexity of college life, an increased and healthy skepticism of the results of narrowly focused "one-shot" studies, and a deepened conviction that one must replicate extensively. In our opinion, many of the findings which pass for facts in the social science literature are not established and, from our experience with successive samples, may never be duplicated.

Before turning to the content of the questionnaire, we have one final comment on the use of the questionnaire as a device for collecting data from college students. Our comment is simply that it is not surprising that college students will answer questionnaires, even such long ones, at a much higher rate than normally encountered in questionnaire studies. It is the nature of people to be interested in themselves and, as observed earlier, particularly during the college years where there occurs a great deal of self-examination. In addition, most college students are interested in education as a general social phenomenon and in educational research if only because so much of their daily life is intimately involved in the educational process. The natural interest in one's self and in one's environment thus tends to favor participation in research focused on the student and his education. Where the campus culture encourages an attitude of responsibility and service to others, as does Yale's, this response tendency is further strengthened.

2. *Questionnaire Content*

Since our study was purposely a multidimensional one, its content can be classified in many different ways. Thus, any description of that content must be somewhat arbitrary in the way it groups individual questions and in the terms it uses to describe those groupings. For present descriptive purposes we have found it most convenient to classify the content in the following manner.

a. For *pre-college background* we wanted an index of the socio-economic position of the subject's family. Accordingly, we asked questions about the total annual family income and marital status of parents as well as the occupation, country of birth, amount of education, and religious preference and participation of both mother and father. For the subject we wanted to know his perception of the social class position of his family, his position in the sibling structure of the family, the types of secondary schools he attended, and the level of his academic aptitude. For the latter we used the verbal score from the Scholastic Aptitude Test and his "predicted average"—a weighted combination of College Board scores and secondary school record (3). These scores were obtained from routine university records rather than from the subject. We also asked the subject for his religious preference and participation, his home state, and country of birth.

b. For *in-college data* to place the student in a general way *vis-à-vis* other students, we ascertained his class in college, school of enrollment (liberal arts or engineering), major field of study, financial aid status (scholarship, loan, student employment), current and cumulative academic average, and participation in various types of athletic and extracurricular activities. We also determined his residential college (dormitory), rooming situation (single or roommates), marital or dating status, amount of spending money available, and whether he had a car at college or easy access to one. (For those who wonder about the last item we might point out that for our campus it is an extremely important one. We are an all-male school within easy driving distance of a great many women's colleges. As our campus humor magazine observes, a bachelor is a fellow who went through Yale without a car.)

c. For *post-college plans and expectations,* we asked for the student's plans for further education, his intended occupation, time when chosen, certainty of choice, expected influence of family in getting his first job, and the number and kinds of different occupations he had seriously considered during college as a possible lifetime work. In addition, we asked for his rating of the importance of eleven different attributes of an "ideal job or profession" for himself. These ranged from a chance to use one's special abilities to a chance to be of service to others. We further asked for anticipated earnings one and ten years after graduation, expected social class position at time of the twenty-fifth reunion, and his ranking of various areas of adult life in terms of the relative satisfaction he hoped to derive from each.

d. The remainder of the questions, which still constituted the bulk of the questionnaire, were centered largely on the student's in-college experience. Under the heading of "*satisfaction*" we asked juniors and seniors for their rankings of seven more or less common denominators of college life at Yale (freshman courses, friends, extracurricular activities, major courses, etc.). Rankings were in terms of the relative contribution made by each to the student's over-all development and preparation for life after graduation and, separately, for the amount of enjoyment derived from each.

To get summary measures of the net satisfaction of the student with his college career we tried both indirect and direct approaches. As an indirect or "barometer" measure of satisfaction a series of questions was asked about the student's emotional reaction to his total experience: would he do it over again, had he considered leaving, would he have enjoyed or profited more from a different type of school, and so forth. More directly, all students were asked how satisfied they were with their current and cumulative academic and non-academic experience. Here and in many other questions previous research had taught us to distinguish carefully between the academic and non-academic (friends, activities, fraternities, etc.) aspects of the college experience. Students vary markedly with respect to the importance they ascribe to each, the experiences they have in each, and in their reactions to those experiences. Measures which fail to differentiate these distinct and important phases of college life may conceal more than they reveal.

To get at the mechanics of the interaction process which was hypothesized to account for the level of the student's satisfaction, we asked a series of questions focused first on the individual dealing with the environment and then on the environment acting on the individual. For the former, we asked for self-ratings of success and difficulty in achieving academic and non-academic goals. We asked for unfulfilled expectations as well as for unexpected benefits in both academic and non-academic terms. Finally, we asked for ratings of twenty-one different "needs" in terms of how much they had been satisfied. These needs were a mixture of academic, social, and health considerations and consisted of such items as sleep, exercise, friends, dates, examinations, vocational preparation, contact with the faculty, guidance in course selection, and so forth. From the standpoint of the environment, the student was asked for self-ratings of success in adjusting to academic and non-academic demands made of him and for the difficulty in effecting that adjustment. He was asked to specify the point in time at which he felt he had achieved a reasonable adjustment academically and non-academically. Finally, we asked for his ratings of success in living up to the academic and non-academic expectations of him held by his father and his mother. Many of the questions pertaining to the interaction process were asked in such a way as to measure the extent of success or difficulty as well as to specify the content of source of that success or difficulty.

e. For *academic achievement*, we already had measures of ability and performance from which an achievement typology could be constructed. The measure of relative achievement type we used was a scale showing the discrepancy between the student's achieved average and the mean average obtained by those with the same predicted average when undergraduate class, type of secondary school attended, and type of study program (engineering/liberal arts) were controlled. As a supplement to this objective index of achievement, we asked for the student's self-rating of his achievement in relation to his own ability. The remaining questions focused on variables believed to be related to achievement. These centered largely on the topics of study habits and time use and included questions about estimated reading speed, note-taking, class-cutting, exam anxiety, and similar topics. In addition a whole series of "motivational questions" were asked. They began with the transition from secondary school to college and asked the student for the reasons for choosing Yale; how many other students were known at entrance; how clear an idea was held of what academic and non-academic life at Yale was like; how clear an idea he had of what *he* wanted to do academically and extracurricularly; and how interested he was in distinguishing himself in each area. Further questions probed his ideas on the purpose of college and his perceptions of how his parents, friends, and most other Yale students felt about the basic purpose of college. He was asked for his beliefs about the relative importance of academic and extracurricular achievement as predictors of success in adult life. The relative pressures from both peers and parents (get good marks, join fraternity, drink etc.) to which he was exposed were also examined as was the over-all importance to him and his parents of his "getting ahead" in the world.

The fact that many of these motivational questions embraced non-academic as well as academic considerations reflects our belief that academic experience must be assessed in relation to other facets of the college experience. Although a college is basically an academic institution and the academic adjustment of the student is crucial if he is to survive in the environment, the fact remains that the college experience is an intricate mixture of both academic and non-academic factors and should be studied with this in mind. One cannot meaningfully assess one factor in isolation from others. We might further observe that if a college's admissions policy is sufficiently rigorous, practically all students admitted are capable of surviving academically. Cases of academic failure are thus usually attributable not to any inherent lack of ability but rather to other "non-academic" factors.

f. To pursue our interest in *student problems* we asked a series of questions focused on difficulties in interpersonal relations that the student had with peers of both sexes, both parents, and "most older people." We inquired about the extent of discomfort engendered by such conditions as headaches, loneliness, insomnia, and indigestion and about the amount of functional disturbance caused by any problem the student might have. By functional disturbance we mean adverse effects of personal problems on important areas of student functioning: studying, sex life, recreation, athletic participation, and so forth.

As indicated, our questions asked students whether they had experienced specified problems or specified effects from any problem. We also were interested in problems which we had not specified. We therefore requested our subjects to describe any current personal problems, the amount of discomfort from such, where they went for help, if anywhere, and how much actual help they thought they had received from whomever they consulted. Because of our special interest in the mental hygiene clinic, we asked those with problems who had not consulted the clinic for their reasons in not doing so. Finally, we asked for self-ratings of physical and mental health in comparison with other students of the same age and sex.

g. Two other areas were the focus of special attention. The first was the *self-concept* of the student and was approached through an adjective check-list of one hundred and sixty items. Here we were essentially interested in what terms students would accept and reject in reference to themselves and in the differences between various groups at a single point in time or at different points in time. With the check-list we hoped to answer such questions as: do freshman under-achievers describe themselves in the same way as senior under-achievers? What are the similarities and differences in self-description between such groups as those who use the psychiatric clinic, those who are satisfied with college and those beset with personal problems.

h. The final area was that of *behavior*. Our approach here was to construct a Reported Behavior Inventory of two hundred and twenty behavioral items and for each to have the student indicate whether he had done what the item described during the school year or, if not, at any previous time. The response options yielded indices of both current and cumulative behavior for single items and groups of items. The items used cross-cut all phases of the college experience and, when grouped, yielded summary measures of the following topics: drinking experience, health, contact with faculty, SPIA (self-propelled intellectual activity), study habits, academic standing, relationship with parents, social experience with girls, work experience, nonathletic active recreation, athletic spectator interest, spectator interest in the arts, athletic activity, heterosexual experience, involvement with college, successful male peer relationships, and moral behavior. Other summary measures could be constructed. These were simply the ones we chose to use. As mentioned earlier, our interest in the behavioral items was primarily a methodological one. We hoped to document our view of the college experience with objective behavioral data as well as by the more subjective data yielded by "mood and attitude" questions.

3. *Question Types*

Although different types of questions were used, most were of the pre-coded variety where the respondent is asked to check the one answer which best approximates his own answer. Others, as implied above, asked the student to rate and/or rank various items according to a definite set of instructions. Still others were of the open-end or write-in variety where the student could express himself in his own terms.

In constructing the questionnaire we tried to group questions by content and to vary the form of the questions to avoid the monotony of similar format. We tried to maintain a balance between questions which would yield answers in comparable form from subject to subject and those which would allow the student free rein in self-expression. Our favorite kind of question was a combination of the two. Here the student was asked to select one of the answers provided and then to explain his answer in his own terms. Aside from the comfort provided the student by such freedom of expression, such questions are useful in a variety of ways. Their answers help illuminate the meaning of statistical results, particularly by indicating the diversity of experience and feeling often concealed by statistics. They provide clues for further research by calling attention to ambiguities in the wording of the question or dimensions of the question overlooked or unintended by the investigator. Where one is involved in a continuous research program, such answers often yield factors which can be quantified and presented as pre-coded response options in subsequent studies. Indeed, many of our pre-coded questions were originally of the open-end variety and had slowly evolved from analysis of open-end answers. Finally, when communicating results, they are an immense help for illustrative and explanatory purposes.

4. *The Sample*

The questionnaire was administered to a random sample of one third of all undergraduates *plus* all those students seen in the mental hygiene clinic who had not been drawn at random. We chose a *random* sample because it would provide a better base for testing and refining our theoretical interests than would a self-selected group. It seemed to us that a model of the college experience should be based on a representative cross-section of all undergraduates rather than restricted to some special group such as those who use the psychiatric clinic. Although on our campus the clinic case load is remarkably representative of the undergraduate body with respect to *objective* indices (type of secondary schooling, grades, fraternity membership, etc.), it is decidedly not representative of the campus with respect to *subjective* indices (attitudes toward self and college, etc.). The basic sample then was the random sample. For special administrative and theoretical reasons, we also included in the study other clinic patients not drawn at random. By removing patients from the random sample and grouping them with the "other patients" we could contrast the entire case load with a cross-section of non-patients. Such a procedure would permit ascertaining clinic use-rates among various segments of the campus as well as accentuate the differences between users and non-users. From the clinic's standpoint, findings would cast in clearer relief the kinds of students and problems with which the clinic was and was not dealing. From the standpoint of the study, intensive study of the case load would yield valuable data on the "problem side" of college, as mentioned earlier. When such special analyses were completed, the patients drawn in the random sample could be removed from the case load and returned to the random sample, thus leaving us with a representative cross-section of undergraduates among whom we could pursue our major theoretical interests. Technically, this procedure is merely oversampling for special purposes but it is highly recommended to those who have special interest in a service agency of the college or who have special interest in a particular sub-group of the student population. It is particularly advantageous where the sub-group of interest is such a small fraction of the total student population that random sampling would produce too few cases to analyze statistically.

The questionnaire was administered to a sample of *all* under-graduate classes to permit the study of change on a cross-sectional basis, as described earlier. It was administered to *one-third* of each class to allow for the shrinkage in numbers which occurs during data analyses when one makes internal comparisons, constructs typologies, and tries to control a variety of variables.

Each questionnaire was marked with an identifying code number and mailed to the subject with a covering letter and a stamped, addressed return envelope. The letter explained the purpose of the study, the method of sampling, and the function of the code number, although in general terms. It requested the student's participation, assured him of confidential treatment of his responses, and invited his comments and suggestions either in person or in writing. Our office phone numbers and locations were contained in the letter.

5. *Code Numbers*

We used identifying code numbers for several reasons. Having a code number rather than a name on a questionnaire reduced the possibility of the questionnaire being associated with the respondent by a clerk or student employee who might know the respondent. The sample list with identifying code numbers was always kept under close security.

We needed to know the respondent's identity in order to do short-term followups to reduce the non-response rate; to contrast at a later date respondents with non-respondents for possible response bias; and to relate the student's answers to other important data which were either not available to him (e.g., his predicted average) or, if theoretically known by him, could not be reliably provided (e.g., his College Board scores). In addition, identifying respondents would permit maximum use of the data where individuals were subjects in other studies being conducted by the clinic or had attributes of special interest to some other office of the university. In the latter instance, since we were working with large numbers of randomly selected students, we were often in a position to select and contrast groups from lists provided by other offices and to provide them additional data on their groups. In such instances identity of individual subjects was never revealed. Reports of analyses were always provided other offices on a group statistical basis. Finally, the use of code numbers permitted the possibility of subsequent studies of the same subjects later in their college careers or as alumni. The latter is an extremely important consideration, for, in our opinion, answers to many questions about the over-all effect of college must come from studies of alumni. Such studies can not be very effective without an adequate in-college base line of data collected from individuals *when they were* undergraduates. Most alumni studies are hampered by the lack of such a base-line and, lacking such a base-line, by the difficulty if not impossibility of trying to reconstruct the undergraduate experience from either official university records or from the memories of alumni (7). In brief, the code number immeasurably increases the usefulness of the data.

It was our conviction that since we needed to identify respondents we should do so openly rather than to conceal identifying symbols from the respondent. We feel strongly that "honesty is the best policy" if one is to establish the rapport with subjects necessary to insure their serious participation in research. This is particularly important if one is involved in a continuing research program, for the reputation, deserved or not, of the agency doing the research is an essential ingredient of its research effectiveness.

From the standpoint of honesty, the use of identifying code numbers necessitates the formulation of a definite policy with respect to use of the data provided, the communication of that policy to the subject, and the rigid adherence to it in actual practice. It was essential that we keep our promise to the student that his individual responses would not be available to any other department of the university or outside agency and that he believe that we would keep it. It is particularly important where one asks students to provide extremely personal data (e.g., sexual practices) or to reveal behavior which, if known to the college, would be grounds for expulsion (e.g., cheating, stealing, etc.). In brief, public relations, if not *the* most important, is one of the most important aspects of successful research where the research depends on data that must be provided by the subjects voluntarily.

While research is often threatening to the individual student, it may be even more threatening to the college administration, particularly where it is concerned, however indirectly, with the impact of various educational practices. Here again there is a need for a mutual respect between researcher and administrative official as well as for a clear set of ground rules with reference to the use of data. It is our belief that it is the researcher's responsibility to make available to the college administration all research results of interest to the administration but without jeopardizing the confidential relationship of researcher and student. He also has the special responsibility of being extremely judicious in the release of specific research findings which may reflect upon the college. This is particularly important where the findings are from "one-shot" studies and have not been thoroughly replicated and where they may reflect adversely upon the college. In brief, *when in doubt,* we feel one should not release results publicly but rather do so privately and then only to responsible college officials and responsible professional colleagues.

D. ANALYSIS

1. *Response Bias*

Eight hundred and thirty-three students or 65 per cent of those solicited in the random sample responded. This response rate, though adequate, was disappointingly low. Our previous questionnaire studies at Yale had obtained response rates ranging between seventy-five and eighty-five per cent. The length of the questionnaire for the present study did not seem to be an important factor in depressing the response rate since we previously had obtained higher response rates with equally lengthy questionnaires. Two factors, however, did seem operative. First, we asked for data of a considerably more personal nature than we had previously. Second, two other large questionnaire studies had been conducted on campus during the same year by other agencies or individuals. Checking the response tendencies of those drawn in more than one of these three studies (something, incidentally, that couldn't be done in the absence of code numbers) revealed that while there is a tendency for students to be consistent in their responsiveness there is also a breaking point. If students answer one questionnaire, they tend to answer other questionnaires; but, as requests accumulate, the probability of answering the latest request seems to diminish. Our questionnaire was administered as close to the end of the academic year as possible and, as the last one, proved to be the "point of no return" for many of those who had also been drawn in the other samples.

To check further on possible sources of bias we compared respondents and non-respondents on a series of characteristics, the data for which had been collected from routine university records. These were: use of psychiatric clinic, number of Yale relatives, SAT-Verbal score, home state by region, academic achievement type, residential college, class in college, types of secondary schools attended, predicted average, and current and cumulative achieved average. By chi-square test the only significant differences found on these items were that the respondents were somewhat more likely to have higher predicted averages ($p < .01$) and higher achieved averages, both current ($p < .05$) and cumulative ($p < .01$). There was also a slight response bias with respect to undergraduate class ($p < .02$), arising from the tendency of sophomores to over-respond and seniors to under-respond.

For two reasons, we made no effort to correct for the slight response bias observed. First, any correction procedure is neccessarily arbitrary and may introduce more bias than already exists. Second, even if one could correct perfectly for bias noted on *objective* indices, one is still left with a *subjective* bias which is inherent in the fact some students did not respond. We can say nothing definitive about the psychological reasons underlying respondent status or about the way in which such reasons may affect the answers given by those who do respond. We therefore prefer to report detectable bias without manipulation of the data.

2. *Control Variables*

Once completed and returned the questionnaires were coded and the data in coded form punched onto IBM cards. Since most of the analysis contemplated would be done on an IBM counter-sorter no special restrictions were placed on the coding format other than the stipulation that each column should be single-punched. For those portions of the data, such as the Reported Behavior Inventory, where we anticipated using the IBM 650 Electronic Calculator, we were careful to use only those symbols which the machine could handle. It is our general recommendation that if the possibility of using electronic calculators exists the researcher should consult with an expert on the machine before he allocates data to specific columns and decides on the coding symbols to be used.

Because of the length of the questionnaire ten cards were needed for each respondent. On each card was punched the subject's identity number, the card number, a number designating the study (an extremely useful device where other studies are being done by the same office) and a series of "control variables." These were variables likely to be used in one or both of two ways: as a major "sorting variable" or as an "index item" which denoted the subject's position on an important facet of the college experience. The sorting variables were measures such as satisfaction or achievement ratings against which we anticipated cross-tabulating all other data on the questionnaire. Simultaneously they served as index items whereby other "non-control" items on any of the IBM cards could be quickly cross-tabulated against what we considered to be the key attributes of the individual. Thus an item such as "being bothered by insomnia" could be quickly checked out against a series of measures summarizing most of the major facets of the student's pre-college background and in-college experience. By the same token any of the latter items could be the subject of intensive study.

The particular control variables we used were: sample status (random or not), clinic use status (patient or not and, if seen, referral source and number of times seen), school and degree for which enrolled, number of hours spent on studying during an average week, the same for extracurricular activities, alumni son, social class position of family, financial aid status (scholarship, loan student employment), predicted average, type secondary school attended, achieved average (cumulative), and self-ratings of physical and mental health, current and cumulative academic satisfaction, and current and cumulative non-academic satisfaction. Also used for controls were five summary measures, each of which subsumed several individual questions on a given topic. These five measures were: a satisfaction score, a study habits score, a score denoting the number of problems and degree of discomfort from them, a typology depicting presence or absence of problems and problem-solving action, and the measure of under-normal-over academic achievement.

3. *Analytical Objectives*

Our over-all analytical objectives were twofold. First, we wanted a composite portrait of the random sample based on their response *as a group* to every item in the questionnaire. Second, we wanted the same composite portrait of major sub-groups within the sample. It is our belief that it is impossible to obtain an adequate non-college control group for college students and that if one is interested in the ultimate question of the effect of college the most rewarding focus for inquiry is on the question of the *differential impact* of a college on groups within its student population. More broadly, we believe the most rewarding line of inquiry is the differential effect of different kinds of colleges on different kinds of students. Accordingly, within our own institution, we were interested in differences between graduates of private and public secondary schools, between overachievers, and underachievers, and so forth. As already mentioned, the major groups in which we were interested were those which could be identified from the control variables.

Permeating all our analytical procedures has been the natural interest of the researcher in developing ways of measuring the subject matter that are both meaningful and accurate. Accordingly, we have been continuously seeking to develop new and better measures of our variables, to refine old measures, and to experiment with ways of condensing and summarizing data.

4. *Procedures*

The most exhaustive technique for determining the basic structure of a mass of data is, of course, to cross-tabulate the answers to each question with those of every other question. This procedure is expensive and time-consuming even where high speed electronic calculators are available. It is also often wasteful of time, money, and energy. We have preferred to be selective in our approach and to use a modified form of cluster analysis. Within discrete areas such as pre-college background data, specific needs, study habit items, satisfaction mood questions, and behavior items, we have cross-tabulated extensively. We have then selected indices of the separate areas and cross-tabluated them. Such a procedure yields an overview of the structure or topography of the data in terms of those features in which we are most interested.

Once we determine the pattern of inter-relationships among the major variables, we then select single topics or variables and use them much like mountain tops as a point from which to survey the surrounding territory. In effect, we ask: what does the college experience look like from the vantage point of academic achievement, . . . of satisfaction, and so forth. Here we first do a massive screening by separating our "sorting variable" into meaningful categories of sufficient size to permit statistical testing and then cross-tabulating these categories against all other data.

From this point on, the direction and kind of analysis is dictated by the initial results. Generally, it consists of progressive refinement of results through assessing the relative importance of variables found to be related to the topic under consideration. For example, on a first run, satisfaction proves to be related to pre-college and in-college variables. The research question then would become whether the in-college correlates are merely a reflection of pre-college correlates or have some importance in their own right. By controlling pre-college correlates we can then proceed to determine what aspects of the college experience are related to satisfaction *regardless of pre-college background.*

In general, this has been our analytical blueprint. As various parts of it are completed, we expect to report the results in a series of papers as we have already done for a few topics (11-14). We expect to focus largely on the topics illustrated by our control variables with occasional special studies of other aspects of the data.

E. SUMMARY

In this paper we have described a large study of the college experience based on self-reports of students. The description has been in terms of the scope, content, methodology, and rationale for the study which was designed to yield a multi-dimensional portrait of the college experience at a single point in time.

The study was rooted in our clinical and research experience over the past decade as staff members of a college mental hygiene clinic and in our conviction that the college experience is an important area of research which, until recently, has been largely neglected. Believing that college students could and should be studied as *students* (young adults in a specific educational environment), we undertook to study as many different aspects of their experience as possible at the same point in time. We were particularly interested in the topics of satisfaction, academic achievement, student problems, behavior, and changes and in their respective correlates and interrelationships.

For our basic research instrument we used a lengthy, self-administered mailed questionnaire which was tailored to our research interests and to the realities of our campus. For our basic sample we used a representative cross-section (one-third) of the undergraduate body and supplemented this with over-sampling of those students who had used the college mental hygiene service. Analysis of the responses of the eight hundred and thirty-three participants (65 per cent response) in the basic sample has been and will be on two levels: (1) the group as a whole; (2) major sub-groups.

We have presented the description of the study at this time and in this way both as a framework for forthcoming reports on various aspects of the study and as a means of communicating our particular views of the importance of studying college students as college students.

REFERENCES

1. BIDWELL, C. E. The American college student and student personality. New York: Social Science Research Council, Committee on Personality Development in Youth, 1960.
2. BIRNEY, R. C., COPLIN, H. R., & GROSE, R. F. The Class of 1959 at Amherst College. Amherst: Amherst College, Committee on Guidance and Counseling.
3. CRAWFORD, A. B., & BURNHAM, P. S. Forecasting college achievement. New Haven: Yale Univ. Press, 1946.
4. DAVIE, J. S., & HARE, A. P. Button-down collar culture. *Hum. Organ.*, 1956, **14**, 13-20.
5. DAVIE, J. S. Satisfaction and the college experience. *In* B. M. Wedge (*Ed.*), *Psychosocial Problems of College Men*. New Haven: Yale Univ. Press, 1958. Pp. 15-44.
6. ———. Who uses a college mental hygiene clinic? *In* B. M. Wedge (*Ed.*), *Psychosocial Problems of College Men*. New Haven: Yale Univ. Press, 1958.
7. ———. Old Blues—New Blues. *In Yale 1945W* New Haven: private printing, 1960.
8. GOLDSEN, R. K., ROSENBURG, M., WILLIAMS, R. M., JR., & SUCHMAN, E. A. What college students think. Princeton: D. Van Nostrand Co., Inc., 1960.
9. NEWCOMB, T. M. Personality and social change: attitude formation in a student community. New York: Dryden Press, 1943.

10. RUST, R. M. Personality and academic achievement: A questionnaire approach. *In* B. M. Wedge (*Ed.*), *Psychosocial Problems of College Men.* New Haven: Yale Univ. Press, 1958.

11. ———. Epidemiology of mental health in college. *J. of Psychol.,* 1960, **49**, 235-248.

12. RUST, R. M., & DAVIE, J. S. The personal problems of college students. *Mental Hygiene,* 1961, **45**, 247-257.

13. ———. Some correlates of self-propelled intellectual activity. A paper read at the annual meeting of the American Psychological Association in New York City, September, 1958.

14. ———. Behavioral change in college. A paper read at the annual meeting of the Eastern Psychological Association in New York City, April, 1960.

15. SANFORD, N. (*Ed.*), Personality development during the college years. *J. Soc. Issues,* 1956, **12**, 3-70.

16. WEDGE, B. M., & DAVIE, J. S. The psychological position of the college man. *In* B. M. Wedge (*Ed.*), *Psychosocial Problems of College Men.* New Haven: Yale Univ. Press, 1958.

Division of Student Mental Hygiene
Department of Health
Yale University
New Haven, Connecticut

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 183-191.

AUTONOMIC EFFECTS ON ESTIMATES OF TIME: EVIDENCE FOR A PHYSIOLOGICAL CORRELATE OF TEMPORAL EXPERIENCE*

Psychology and Environmental Medicine Divisions,
U.S. Army Medical Research Laboratory

GLENN R. HAWKES, ROBERT J. T. JOY, AND WAYNE O. EVANS

A. INTRODUCTION

The search for physiological correlates of performance when judging intervals of time has yielded apparently conflicting results. Hoagland (8) reported that an influenza patient utilizing the method of production (cf. 1) indicated the passage of one minute subjective time in considerably less than one minute physical time. Hoagland attributed this result to increased bodily temperature or the reaction rate of a chemical pacemaker in the CNS. On the other hand, Schaefer and Gilliland (9) were unable to relate time judgments to such physiological indices as pulse rate, respiration rate, and blood pressure. Wallace and Rabin (10) reviewed suggestions by a number of investigators on the internal factors in time estimation, and concluded that hypotheses now available are inadequate (also see 11).

Experimental alteration of the activity of the autonomic nervous system by administration of drugs has been shown to affect time judgments (N.B. 4). Results in many studies have been inconsistent, however, and typically the dosage levels have been far in excess of normal therapeutic usage (5, 10). One of the aims of the present study was to determine the effect of the usual doses of autonomically active drugs upon judgments of time.

By utilizing different methods of time estimation, the present study also attempted to ascertain physiological correlates of different conditions of temporal judgment as drugs produce variations in autonomic activity. Frankenhaeuser (4) notes that time estimation by the method of production apparently is based upon internal (physiological) events. On the other hand, judgments by the reproduction method are mostly under the control of the external environment. These two methods, however, do not result in significantly different judgments of auditory, visual, or cutaneous stimulus

---

* Received in the Editorial Office on September 22, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

duration for normal (non-drugged) subjects (7). When autonomic activity is altered by drugs, then it may be predicted that judgments with the production method will be correlated with variations of such internal events as heart rate, respiration rate, etc., whereas reproduction judgments will not be correlated with such internal activity.

In order to permit comparison with responses of untreated subjects, the stimuli and durations were the same as in a previous study (7). Judgments of duration were made for intervals of white noise, white light, and random-frequency electrical cutaneous stimuli. The methods of production and reproduction were used with a duration range of 0.5 to 4.0 seconds. Records also were taken of heart rate, skin temperature, respiration rate, and blood pressure.

B. METHOD

Seven subjects each participated in seven sessions. In an individual session one of seven treatment conditions was used (in a counterbalanced order). Table 1 lists the drugs and doses and their predominant action.

TABLE 1

Drug	Dose	Predominate action
1. Methamphetamine HCL	2.5 mg.	central sympathetic stimulation
2. D—amphetamine SO_4	5.0 mg.	general sympathetic stimulation
3. Phenylpropanolamine HCL	50 mg.	peripheral sympathetic stimulation
4. Chloropromazine HCL	25 mg.	general sympathetic blocking
5. Mecamylamine HCL	2.5 mg.	peripheral ganglionic blocking
6. Placebo (Lactose)	50 mg.	control
7. No drug		control

The drugs used in this study were chosen for their predominant effects on peripheral or central aspects of the autonomic nervous system. Dosage was consistent with standard therapeutic usage (cf. 6). All drugs were administered in identical capsules taken orally with water. A double-blind procedure was used so that neither *S* nor *E* was aware of the specific drug administered in a particular session.

After a 16-hour fast, the subject lay in a supine position on a canvas web couch wearing only shorts. The experimental chamber was maintained at a temperature of 28.3° C, SD 4.2° C (sampled every 20 minutes for all sessions). These values are in the range of critical temperatures for nude man, and ensure that there was no ambient thermal stress (3). The chamber was sound-treated, and the subject was alone throughout the testing period. Skin temperature was sensed by four copper constantan thermocouples attached to the anterior chest, and was recorded on a Brown potentiometer with a recording cycle of 0.5 of a second. Systolic blood pressure was measured with a modification of Zuidema's method (12) using a photoelectric plethysmograph to sense the pulse in the index finger. Respiratory rate was obtained using a mask-mounted strain gage pneumotachograph (2). The heart rate was determined from an electrocardiagram. These last three variables were recorded on two Sanborn Twin-Viso recorders.

With the subject in position with leads attached, basal physiological measurements were secured while he remained at rest for one hour. He then was given the appropriate drug and rested in quiet for two additional hours. Throughout this 3 hour period, (1) skin temperature was recorded every 0.5 of a second, (2) systolic blood pressure was recorded every 60.0 seconds, and (3) heart rate and respiratory rate were continuously monitored.

After this 3 hour period, the physiological recording devices were removed, and subjects then made judgments of stimulus duration with three modalities, two methods, and four intervals. Physiological recording coincident with time estimation was not possible due to electrical interference. The intervals were 0.5, 1.0, 2.0, and 4.0 seconds, in a different random order for each session. Responses were made by switching on and off the stimulus using the methods of production and reproduction. The stimuli were (1) white noise presented binaurally through earphones, (2) white light presented through an 11 × 17 inch viewing window, or (3) a random-frequency electrical signal presented through a 12 mm. diameter electrode resting on the index finger pad. Order of presentation for method and modality was counterbalanced. Intensity of the electrical cutaneous stimulus was 160 per cent of the amount of current necessary to reach absolute threshold. The intensity of the visual or auditory stimuli was determined by having the subject equate subjective intensity to the apparent intensity of the cutaneous stimulus. (For a fuller description of the apparatus and procedure used in the duration judgments, see 7.)

C. RESULTS

The data for the four physiological measures are reported as deviations from the base line. The average value of each physiological parameter for the 24 minutes just prior to the time estimations was subtracted from its average value for the 30.0 minutes just prior to the administration of the drug. The mean differences for all subjects for each parameter and each drug condition are summarized in Table 2. Tables 3, 4, 5, and 6 present the results of analyses of variance for each of the physiological indices,

TABLE 2

MEAN DIFFERENCES OF PHYSIOLOGICAL INDICES FROM BASE LINE

	Amphetamine	Methamphetamine	Phenylpropanolamine	Chloropromazine	Mecamylamine	Placebo	Control
Heart rate in beats/min.	2.6	.3	1.3	—1.6	—2.3	—1.7	—3.0
Blood pressure in mm. Hg	3	2	14	2	—6.4	—2	0
Skin temperature in ° C	.3	—.3	0	—.3	—.1	—.1	0
Respiration in breaths/min.	2.0	0.9	2.8	0	—0.5	—0.1	0.7

TABLE 3

ANALYSIS OF VARIANCE OF THE EFFECTS OF DRUGS ON RESPIRATORY RATE

Source	SS	df	MS	F
Total	40528	48		
Drugs	25900	6	4317	21.59[a]
Subjects	7420	6	1237	6.19[a]
Drugs × Subjects	7208	36	200	

[a] $P < .01$.

TABLE 4

ANALYSIS OF VARIANCE OF THE EFFECT OF DRUGS ON HEART RATE

Source	SS	df	MS	F
Total	218356	48		
Drugs	34906	6	5818	1.64
Subjects	55591	6	9265	2.61[a]
Drugs × Subjects	127859	36	3552	

[a] $P < .05$.

TABLE 5

ANALYSIS OF VARIANCE OF THE EFFECT OF DRUGS ON TEMPERATURE

Source	SS	df	MS	F
Total	521109	48		
Drugs	137479	6	22913	2.83[a]
Subjects	92548	6	15425	1.91
Drugs × Subjects	291082	36	8086	

[a] $P < .05$.

TABLE 6

ANALYSIS OF VARIANCE OF THE EFFECT OF DRUGS ON BLOOD PRESSURE

Source	SS	df	MS	F
Total	103013267	48		
Drugs	18100219	6	3016703	1.42
Subjects	18474648	6	3079108	1.45
Drugs × Subjects	76438400	36	212389	

and Table 7 summarizes the analysis of variance of the time estimation data.

Table 8 presents Pearson Product-Moment correlation coefficients of particular interest for this paper. These were selected from a matrix of all possible comparisons of the 28 dependent variables as the most demonstrative of the hypothesis. The correlations were obtained by considering the dependent variable across all drugs and subjects.

D. DISCUSSION AND CONCLUSIONS

The physiological changes induced in the present group of subjects by the treatment conditions indicates that usual therapeutic doses of the drugs can

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE OF TIME ESTIMATES

Source of variance	df	MS	F	P
Method (A)	1	5.3044	—	
Modality (B)	2	1.9205	5.68	< .05
Intervals (C)	3	716.5527	381.14	< .01
Drugs (D)	6	.7637	—	
Subjects (E)	6	19.7542	—	
A × B	2	.8554	6.29	< .01
A × C	3	7.7736	4.14	< .05
A × D	6	.3131	—	
A × E	6	10.9904	5.85	< .01
B × C	6	.0162	—	
B × D	12	.2051	—	
B × E	12	.3380	2.49	< .01
C × D	18	.2681	—	
C × E	18	2.7409	—	
D × E	36	.8864	2.04	< .05
A × B × C	6	.3393	4.71	< .01
A × B × D	12	.1565	—	
A × B × E	12	.1153	—	
A × C × D	18	.0798	—	
A × C × E	18	1.8792	12.86	< .01
A × D × E	36	.4354	2.98	< .01
B × C × D	36	.0768	—	
B × C × E	36	.0740	—	
B × D × E	72	.1814	—	
C × D × E	108	.2601	1.78	< .01
A × B × C × D	36	.0609	—	
A × B × C × E	36	.0469	—	
A × B × D × E	72	.1359	1.89	< .01
A × C × D × E	108	.1462	2.03	< .01
B × C × D × E	216	.0761	—	
A × B × C × D × E	216	.0721	—	

significantly change some aspects of autonomic activity. Of particular interest is the relatively large effect of drugs on respiration rate. In contrast is the failure to find a significant main effect of drugs in the analysis of variance of the time estimation data and the failure to find any meaningful pattern of significant interactions involving drugs. An apparent exception is the interaction of drugs with methods and subjects, which may be interpreted as a drug effect on duration judgments for some subjects when using the production method.

The significant effect of modality of stimulation ($P < .05$) indicates that the sensory channel of stimulation affects the accuracy of time judgments, in agreement with a previous report (7), but no meaningful pattern of significant interactions involving modality is apparent to the writers. The most interesting result of the analysis summarized in Table 7 is the consistent

TABLE 8
PEARSON PRODUCT-MOMENT CORRELATIONS FOR HEART RATE AND RESPIRATION RATE WITH TIME JUDGMENTS

Method	Modality	Interval			
		½	1	2	4
			Heart Rate		
Production	cutaneous	—.26	—.17	—.36[a]	—.37[b]
	audition	—.35[a]	—.33[a]	—.40[b]	—.38[b]
	vision	—.22	—.30[a]	—.44[b]	—.37[b]
Reproduction	cutaneous	—.20	—.19	—.06	—.04
	audition	—.33[a]	—.14	—.04	—.25
	vision	—.26	—.08	—.23	—.08
			Respiration Rate		
Production	cutaneous	—.50[b]	—.38[b]	—.39[b]	—.47[b]
	audition	—.31[a]	—.38[b]	—.39[b]	—.49[b]
	vision	—.35[a]	—.27	—.40[b]	—.34[a]
Reproduction	cutaneous	—.14	+.14	—.12	—.16
	audition	—.02	+.28	+.16	—.08
	vision	+.19	+.29[a]	—.01	—.14

[a] $P < .05$.
[b] $P < .01$.

pattern of significant interactions involving method; particularly noteworthy in this connection is the relatively large *F*-ratio for the interaction of method with intervals and subjects. It may be seen that much of the subject variance is accounted for by the use of more than one method of time estimation.

Judgments of stimulus duration with the method of reproduction did not tend to correlate significantly with changes in autonomically influenced physiological variables (see Table 8), supporting the statement made in the Introduction that such judgments are mostly under the control of the external environment. The significant correlations of variation in heart rate and respiration rate with changes in time estimates also indicates that the method of production is a useful technique for revealing changes in temporal judgments correlated with internal events. Although there was a tendency for both heart rate and respiration rate to correlate significantly with production judgments, it is obvious that respiration rate is the best index of such changes. The fact that the significant correlations are negative also indicates that as autonomically influenced physiological activity speeds up, subjective time also tends to accelerate, with the result that the subject will indicate the passage of a specified period of time in much less than the specified period of physical time. This finding is in agreement with the report of Hoagland (8) reviewed earlier.

Pearson correlation coefficients not presented in this report also indicated that the duration judgments did not correlate significantly with skin temperature or blood pressure. Further, the various physiological measures did not significantly correlate with each other. It was apparent, however, that the judgments of a given interval in general correlated significantly with judgments of the same interval using another modality or method (0.5 of a second auditory production with 0.5 of a second visual production, etc.). It also was apparent that intramodality correlations were in general significant, while intermodality correlations were not consistent. Inasmuch as these correlational results are to be expected, it is of more interest that judgments with the method of production did not tend to correlate with judgments via the reproduction method, lending further support to the notion that use of the production method affords greater opportunity to detect changes in performance correlated with variations of autonomically influenced physiological activity.

The failure of Schaefer and Gilliland (9) to find physiological correlates of temporal experience could have been due to their use of longer intervals than those of the present study, or to their use of the method of verbal estimation (see 7). More probably, however, the discrepancy with the present report is due to the fact that Schaefer and Gilliland did not allow their subjects to rest for a sufficient period prior to the measurements and thus allow for stabilization of the physiological indices. (Changes in autonomic activity in their study were induced by exercising the subjects.) This explanation receives credence by virtue of the extreme variability of the physiological responses of their (four) subjects, as reported in their paper.

It may be concluded that it is possible to obtain correlations of rhythmic physiological activity to temporal estimates if one uses the method of production and allows sufficient time for the physiological measures to stabilize prior to obtaining such estimates. With the method of production, it may be assumed that no external referent was available in the present study, *ergo* the subject based his temporal estimate upon internal (physiological) activity.

E. SUMMARY

The present study investigated the possible correlation of autonomically influenced physiological activity variations with changes in estimation of stimulus duration. The methods of production and reproduction were used with visual, auditory, and electrical cutaneous stimuli. Measurements were taken of chest skin temperature, blood pressure, heart rate, and respiration

rate. Mild variations of these indices were produced by administration of autonomically active drugs at usual clinical dosage levels.

The results indicated that respiration rate and heart rate were correlated with temporal judgments when using the method of production, but judgments did not correlate with blood pressure or skin temperature. Consistent significant correlations were not apparent when using the method of reproduction.

It is concluded that duration judgments with the method of production, in which time estimates are based upon internal events (if external cues are not available), will correlate with respiration rate changes and perhaps with other indices of rhythmic physiological activity such as heart rate. Temporal judgments with the method of reproduction do not significantly correlate with autonomically influenced physiological variables.

REFERENCES

1. BINDRA, D., & WAKSBERG, H. Methods and terminology in studies of time estimation. *Psychol. Bull.,* 1956, **53**, 155-159.
2. DAHLKE, H. E., & WALKOWITZ, W. Diagnosing with strain gages. *Elec. Ind.,* 1959, No. 1.
3. ERICKSON, H., KROG, J., ANDERSON, K. L., & SCHOLANDER, P. F. The critical temperature in naked man. *Acta Physiol. Scand.,* 1956, **37**, 35-39.
4. FRANKENHAEUSER, M. Estimation of Time. An Experimental Study. Stockholm: Almqvist & Wiksell, 1959.
5. GILLILAND, A. R., HOFELD, J., & ECKSTRAND, G. Studies in time perception. *Psychol. Bull.,* 1946, **43**, 162-176.
6. GOODMAN, L. S., & GILMAN, A. The Pharmacological Basis of Therapeutics. New York: MacMillan, 1956.
7. HAWKES, G. R., BAILEY, R. W., & WARM, J. S. Method and modality in judgments of brief stimulus duration. *J. Aud. Res.,* 1961, **1**, 133-144.
8. HOAGLAND, H. Pacemakers in Relation to Aspects of Behavior. New York: MacMillan, 1935.
9. SCHAEFER, V. G., & GILLILAND, A. R. The relation of time estimation to certain physiological changes. *J. Exp. Psychol.,* 1938, **23**, 545-552.
10. WALLACE, M., & RABIN, A. I. Temporal experience. *Psychol. Bull.,* 1960, **57**, 213-236.
11. WOODROW, H. Time perception. *In* Stevens, S. S. (*Ed.*), *Handbook of Experimental Psychology.* New York: Wiley, 1951.
12. ZUIDEMA, G. D., EDELBURG, R., & SALZMAN, E. A. A device for indirect recording of blood pressure. *J. Appl. Physiol.,* 1956, **9**, 132-134.

Civil Aeromedical Research Institute
Federal Aviation Agency
Oklahoma City, Oklahoma

Published as a separate and in *The Journal of Psychology,* 1962, **53,** 193-198.

REINFORCEMENT AS A FUNCTION OF INSTRUCTIONAL SET*[1]

Department of Psychology, Douglass College, Rutgers University

DONALD G. FORGAYS AND HAZEL MOLITOR

A. PROBLEM

There have been a score or more experiments reported recently concerned with the utility as reinforcers of verbal cues ("good," "mmm-hmmm, etc.) or mechanical cues (lights or bells) in the modification of verbal or simple motor behavior (see references 4, 6, 7, 15, 16, 17, 34, 35, 36, 39, 40, 44, 61, 64, 68, 91, 96, 97, 101, 105, and 107 in Krasner, 1958; also Caditz, 1953; Hall, 1960; Terrell and Kenedy, 1957; and others). Results reported in these studies have not been in close agreement as to the efficacy of certain cues as reinforcers. Nearly all researchers report that verbal cues such as the spoken word "good" operate effectively as reinforcers in the operant conditioning of a variety of verbal and a few motor responses. Mixed results have been reported, however, with respect to the effectiveness of mechanical cues in similar situations. In some cases lights or sounds were found to have reinforcement value; in others, little or none.

*S*s in most of the studies referred to are typically provided with minimal instructions as to the nature of the task, that is, the behavior expected of them, and as to the meaning of the verbal or mechanical cue which is to appear. Under the circumstances of this experimental design, it is possible that any sort of experimenter feedback, interpretable by the *S* as such, may operate as a reinforcer, in the operational sense. It is likely, moreover, that verbal cues will be more effective as reinforcers than mechanical cues simply because they are more readily interpretable as experimenter feedback responses. Finally it is to be expected that the nature of the original instructions to the *S* would contribute importantly to the efficacy of the various cues employed as reinforcers. Specifically, one could suggest that a light or bell might operate as a reinforcer just as effectively as a verbal cue

---

* Received in the Editorial Office on September 27, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was supported in part by a research grant to the senior author from the Rutgers University Research Council. A paper based upon the study was reported at the annual meetings of the Eastern Psychological Association in Philadelphia on April 7, 1961.

if the instructional set provided the *S* led to his acceptance of the cue as an experimenter feedback response indicating acceptable or desired behavior.

While there has been long concern in the psychological literature with the influence of set or instructions upon behavior, there has been relatively little research reported of such effects upon the kind of operant conditioned responses of interest here. A few researchers do report the variation of instructional set in verbal operant conditioning designs (Hall, 1960; Sidowski, 1954; and Tatz, 1956). Each reports results indicating an effect of instructional set upon level of conditioning achieved. Their results indicate essentially little difference in conditioning between non-informed *S*s who receive reinforcement (groups somewhat comparable to those of the minimal instruction experiments referred to above) and those *S*s who expect reinforcement and receive it. In none of these experiments is there a systematic variation of type of reinforcing cue and therefore the possible interaction of instructional set and reinforcing cue is left unstudied.

The present experiment examines the influence upon the operant conditioning of a simple motor task of two levels of instructions with several types of reinforcing cues.

B. METHOD

1. *Subjects*

One hundred female college students enrolled in the introductory psychology course at Douglass College, Rutgers University, were used as *S*s.

2. *Apparatus*

Apparatus for the *S* consisted of a small box, 6½″ by 6½″ by 4″ high, on top of which were mounted in a row five small button switches and a red pilot lamp and in back of which was mounted a small code-type buzzer. Pressing of any of these buttons closed a circuit which lighted an individual indicator light for the *E* and also operated a counter. The five indicator lamps and counter were mounted on another small box for the *E* on which were also mounted switches for operating the light and the buzzer for the *S*. *S* and *E* were separated across a table by a plywood screen.

3. *Procedure*

*S*s were assigned at random to the various subgroups of the study and were then tested individually. Half of the *S*s were instructed by the *E* to press one or more of the buttons before them on each of 10 consecutive 30-second trials. This condition is labeled "incomplete instructions." The remaining 50 *S*s were given these same instructions and were told, in addition, that when they pressed in a manner which the *E* had in mind, a distinctive cue would occur; the specific cue to be employed for each *S* was indicated to her, save for the control group which was told only to expect a distinctive cue. This latter general condition is labeled "complete instructions."

Within each of these two main groups of *S*s there were five subgroups of 10 *S*s each. The first was reinforced by the *E* saying the word "good" at the appropriate time; the second, by a one-second flash of the light on the *S*'s box; the third, by a one-second signal of the buzzer; the fourth, by one second of either the light or the buzzer presented approximately equally on a random basis; and the fifth was not reinforced by any of these cues (control groups). A ratio schedule was employed; 10 individual presses of any combination of buttons led to reinforcement of appropriate groups. The *E* recorded total number of button presses per trial directly from the electrical counter. The *E* was a young female adult, a senior honors student in psychology at Douglass College.

C. RESULTS

Mean level of button-press response totaled for the 10 trials for each of the 10 groups of *S*s is presented in Table 1.

TABLE 1
MEAN NUMBER OF PRESSES OF THE FIVE SUBJECT GROUPS FOR EACH OF THE TWO LEVELS OF INSTRUCTION. N = 10 IN EACH GROUP

	Reinforcement				
	"Good"	Light	Sound	Light or Sound	Control
Complete instructions	235.2	298.8	264.9	274.7	468.6
Incomplete instructions	498.6	342.2	498.2	388.1	301.1

As can be seen there, under complete instructions the control group responded at a considerably higher level than the four reinforced groups, all of which responded at approximately the same level. Under incomplete instructions, the control group responded considerably less than the four reinforced groups. In this latter condition, there was more variability of response level among the reinforced groups, some responding close to the control group level and others responding much more highly, at a level equivalent to that of the control group under complete instructions.

A summary of an analysis-of-variance of these data is presented in Table 2. As can be seen there, the between-groups effect is significant at the .05 level of confidence. This effect is further broken down into its three components: effect of instructions, effect of reinforcement conditions, and the interaction of instructions and reinforcement. The main effect of instructions and the interaction effect: instructions by reinforcement were both found to be significant at the .05 level of confidence. The main effect of reinforcement condition was not found to be significant.

TABLE 2
SUMMARY OF THE ANALYSIS-OF-VARIANCE OF THE RESPONSE DATA FOR THE TEN GROUPS OF SUBJECTS

Source	Sum of squares	df	Mean square	F	P
Between groups	892,029.8	9	99,114.4	2.15	$<.05$
Instructions	246,016.0	1	246,016.0	5.34	$<.05$
Reinforcement	75,240.7	4	18,810.2	0.41	N. S.
Interaction	570,773.1	4	142,693.3	3.10	$<.05$
Within groups	4,144,736.0	90	46,052.6		
Total	5,036,765.8	99			

Multiple range testing (Duncan, 1955) of individual comparisons indicates that under complete instructions the control group responded at a significantly higher level than the groups which received as reinforcers the word "good," the buzzer, and the light-buzzer, all at the .05 level of confidence and at a level higher than the group which received light, at the .10 level of confidence. There are no significant differences among those groups which received reinforcing cues under complete instructions. Under incomplete instructions, the groups which received as reinforcers the word "good" or the buzzer responded at a significantly higher level than the control group, at the .05 level of confidence. Again, there are no significant differences among the reinforced groups. Comparing the effectiveness of reinforcement conditions under the two levels of instructions indicates that the groups which received "good" or buzzer responded at a significantly higher level under incomplete than complete instruction, at the .05 level of confidence, but that the control group under complete instructions responded at a significantly higher level than under incomplete instructions, at the .10 level of confidence.

D. DISCUSSION

The results of this study indicate clearly that the efficacy of reinforcement such as that used here is a function of the instructional set provided the *S*. With incomplete instructions, which are most typically employed in related studies, *S*s who received any of the reinforcing cues used responded at a level higher than that of *S*s who received no such cues; two of these four comparisons are statistically significant, but all mean differences are in the same direction. There is some indication that auditory cues are more effective as reinforcers than the visual cue used, since the groups which received the word "good" or the buzzer respond at an appreciably higher level, though not significantly so, than the groups which received light or the light-buzzer combination. It is likely that this difference is due to the discriminability of the cues employed, the medium-loud buzzer or the spoken word being considerably more distinctive than the flashing of the small pilot lamp in a well-lighted room.

With complete instructions, all *S* groups which received reinforcing cues responded at approximately the same level and considerably less than the control condition, three of these four comparisons being statistically significant. Being prewarned as to the nature of the reinforcing cue to appear later had an apparent decremental effect upon level of response for those groups receiving such cues but a rather striking incremental effect upon the control group which expected to receive a reinforcing cue but did not. This latter group responded at a level approximately as high as any group within the present study.

Direct comparison of the effectiveness of the various reinforcing cues across instruction conditions indicates that the relative influence of instructions was largely for the groups which received as reinforcers the word "good" or the buzzer and the groups which received no reinforcing cue at all, with the latter effect being in the opposite direction. The fact that such differences were not found for the light and light-buzzer groups may be a function, again, of the relative discriminability of such cues, especially since light has been found previously to be an effective reinforcing cue, at least under some conditions (Krasner, 1958).

These findings suggest that the set of the *S* is a critical variable in such learning designs. It appears that, if a specific set is not provided the *S* through experimental instructions, a task for the *S* in such a study is to provide his own instructions which he likely does in terms of the feedback responses of the *E*. Under these conditions any distinctive cue will likely operate as a reinforcer.

In light of the fact that mixed results have been reported as to the efficacy of various signals as reinforcers (Krasner, 1958, and others), the present findings suggest a possible source of confusion and a possible reinterpretation of many similar studies which employed incomplete instructions and verbal and/or other sensory cues as reinforcers.

E. SUMMARY

The present experiment investigated the effects of instructional set and type of reinforcement on the modification of an operant button-press response. One hundred female college students were assigned equally to 10 groups. Two of these groups were reinforced with the word "good" spoken by the *E*, two with a light-flash, two with a buzzer sound, two with the light-flash or the buzzer sound, and two received no reinforcing cue. One of each of these pairs of groups was provided instructions to expect a feedback cue after appropriate behavior; the remaining five groups were not provided these instructions. All *S*s were told to press one or more of the five buttons provided one or more times on each of the 10 consecutive 30-second trials. The eight reinforced groups were given the appropriate feedback cue for each combination of 10 button presses.

Results indicate that the efficacy of reinforcement such as that used is a function of the instructional set provided the *S*. With incomplete instructions, *S*s who received any of the reinforcing cues responded at a higher level than control *S*s. With complete instructions, all *S* groups which received reinforcing cues respond at a level considerably less than the control condition.

It appears that if a specific set is not provided the *S* through experimental instructions, a task for a *S* in such a study is to provide his own instructions, which he likely does in terms of the feedback responses of the *E*. Under these conditions any distinctive cue will likely operate as a reinforcer, in the operational sense.

REFERENCES

1. CADITZ, S. B. A study of partial reinforcement of a lever-pressing response in young children. Unpublished M.A. thesis, Univ. of Washington, 1953.
2. DUNCAN, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1-42.
3. HALL, W. E. The effects of set and reinforcement in verbal conditioning. *J. Gen. Psychol.*, 1960, **63**, 239-248.
4. KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148-170.
5. SIDOWSKI, J. B. Influence of awareness of reinforcement on verbal conditioning. *J. Exp. Psychol.*, 1954, **48**, 355-360.
6. TATZ, S. J. Symbolic mediation in "learning without awareness." *Amer. Psychol.*, 1956, **11**, (No. 9), 498.
7. TERRELL, G., & KENNEDY, W. A. Discrimination learning and transposition in children as a function of the nature of the reward. *J. Exp. Psychol.*, 1957, **53**, 257-260.

Department of Psychology
Douglass College
Rutgers University
New Brunswick, New Jersey

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 199-209.

THE PITTSBURGH SCALES OF SOCIAL EXTRAVERSION-INTROVERSION AND EMOTIONALITY*

Department of Psychology, University of Pittsburgh

A. W. BENDIG

A. INTRODUCTION

The development by Eysenck (1959) of the Maudsley Personality Inventory containing 24-item scales of "extraversion" (E) and "neuroticism" (N) has stimulated almost as much research relating these personality dimensions to laboratory measures of human learning as did the previous publication of Taylor's Manifest Anxiety Scale (Taylor, 1953). In developing his scales, Eysenck (1956) administered 261 items from Guilford's previously developed S, D, C, R, G, and A scales to 400 British *S*s (200 men and 200 women). Taking the R ("rhathymia") scale as the best available measure of "extraversion-introversion" and the C ("cycloid disposition") scale as the best measure of "neuroticism," each item was correlated with both the R and C scores separately within each sex group. The 24 items correlating highest with R and lowest with C were selected for the MPI E scale and those correlating highest with C and lowest with R became the MPI N scale. A subsequent factor analysis of the selected 48 MPI items with (presumably) the same group of *S*s confirmed the assignment of the items to the two scales, although several items showed appreciable loadings on the "wrong" factor or weak loadings on the "right" factor.

Subsequent research has indicated that the MPI scales measure two second-order personality factors consisting of intercorrelated clusters of first-order factors. Jensen (1958) noted that the E scale correlated highly with a questionnaire measure of "social introversion," but not with measures of "thinking introversion" or "emotional introversion" and suggested that the E score is a measure of "social extroversion-introversion." Research utilizing the factor validated scores of the Guilford Zimmerman Temperament Survey (Guilford and Zimmerman, 1949, Bendig, in press) has shown the MPI E scale to correlate highly with Guilford's A ("ascendance") and S ("sociability") factors and somewhat less highly with his G ("general activity") and R ("restraint," the mirror-image of Guilford's earlier "rhathymia")

---

* Received in the Editorial Office on September 29, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

factors, while the N scale correlated with Guilford's E ("emotional stability," a mirror-image combination of earlier "cycloid disposition" and "depression"), O ("objectivity"), and, to a lesser extent, S ("sociability"), P ("personal relations"), F ("friendliness"), M ("masculinity") and T ("thoughtfulness" or "thinking introversion") factors (Bendig, in press). Preliminary factor analyses of the MPI items has given evidence that clusters of items measuring first-order factors can be found within the E and N scales (Bendig, 1960b). Other research has shown that the MPI N scale measures a broad second-order "emotionality" (Em) factor that is also measured by other "neuroticism" and "anxiety" scales (Bendig, 1960a), but that other scales loaded with this factor, such as Taylor's MAS, are more contaminated by an "introversion" component than is Eysenck's scale (Bendig, 1957, 1958, 1960a).

The use of the MPI in personality research presents some difficulties. The standard MPI (Eysenck, 1959) requires the *S*s to give trichotomous responses, i.e., yes, question mark, or no, with two, one, and zero weights being assigned to the response options and these weights being summed for the score on each scale. This method of scoring the items requires hand scoring of the answer sheets or, if the test is adapted for machine scoring, two runs of the answer sheets through an IBM test scoring machine for each scale. However, research has shown that dichotomous responses, i.e., yes, no, or true, false, can be used with the MPI items without lowering the reliabilities of the scales and with a considerable saving in scoring labor (Bendig, 1959a, 1959b).

A second practical difficulty is that the MPI items, taken from Guilford's earlier inventories, are written as second person interrogative statements, rather than the first person declarative items characterizing most other personality questionnaires, such as the MMPI and the almost innumerable scales derived from it (such as Taylor's MAS), or the second person declarative items used in the Guilford Zimmerman Temperament Survey. These differences in item styles among the MPI, GZTS, and MMPI inventories present a problem in attempting to integrate scales from the three instruments into new questionnaires for further personality research. The simplest solution would be to change all items to a common style to avoid any possible, albeit unlikely, problems to the *S*s in changing mental set in responding to different items, but such a change could possibly reduce the factorial validity and independence of the scales. Consequently, the changed items require re-validation of the scales through factor analytic procedures.

A third problem with the MPI lies in the content of the items. As noted above, the E scale is heavily loaded with socially oriented items and contains few readily identifiable as "rhathymia" items. Similarly, neither scale includes many identifiable with Guilford's "thinking introversion" factor. This suggests that the MPI scales, derived by item analytic procedures on British *S*s, may be too narrow in item scope for research with American *S*s. It is instructive to note that the MPI scales were not derived through factor analysis, but items were selected through item analysis and factor validated by an *ex post facto* factor analysis of the already selected 48 items (Eysenck, 1956). This factor analysis, presumably performed using the item responses of the same British *S*s used in selecting the items, and two factor analyses using a small sample of 100 American *S*s (Bendig, 1960b) provide the only available evidence as to the factorial validity of the MPI items.

The research reported below was undertaken with the aim of developing reliable and factorily valid scales measuring the two broad second-order factors of "social extraversion-introversion" (SEI) and "emotionality" (Em) for use in research with American college *S*s.

B. PROCEDURE

1. *Items*

An inventory containing 273 items was prepared for administration to the *S*s. The 48 Extraversion and Neuroticism items of the Maudsley Personality Inventory (Eysenck, 1959) were changed from their second person interrogative form to first person declarative and 40 Restraint and Thoughtfulness items from the Guilford Zimmerman Temperament Survey (1949) were similarly converted from second person declarative to first person declarative for inclusion in this questionnaire. The GZTS items used were primarily those previously found to be most highly loaded with Guilford's R and T factors (Bendig, 1961) with the exception that five items (GZTS nos. 27, 77, 82, 193, and 253) appeared to overlap in content with MPI items and were replaced by five other GZTS items (nos. 47, 107, 112, 122, and 223) that had loaded somewhat less highly on the R and T factors, but whose wording appeared not to duplicate any of the other items.

Also included in the inventory were the 50 items of Taylor's Manifest Anxiety Scale (1953) and 20 items taken from the MMPI Social Introversion (Si) scale (Drake, 1946).[1] Included as "buffer" items were 115 items from several experimental hostility and need achievement scales that were not included in the present analyses. All of the scales were randomly intermixed in the questionnaire.

---

[1] The MMPI Si scale items included in the present research were selected from two previous item analyses of selected Si items (reversed in scoring). In the first analysis, 20 Si items were included in an inventory along with Eysenck's E and N scales which was administered to 253 college *S*s of both sexes. Each Si item was

The instructions to the *S*s requested them to respond with dichotomous true-false responses on a separate IBM answer sheet. The responses to the GZTS Restraint and MMPI Social Introversion scales were reversed in scoring from the published keys to provide measures of "rhathymia" and "social extraversion."

2. *Subjects*

The inventory was administered during a regular class session within the first two weeks of each trimester to all undergraduate students enrolled in a one-semester introductory psychology course during the Fall, Winter, and Spring trimesters at this university. There were 480 *S*s (236 men and 244 women) in the Fall, 293 *S*s (179 men and 114 women) during the Winter, and 131 *S*s (87 men and 44 women) during the Spring trimesters. The answer sheets of the *S*s were alphabetized by their surnames within each trimester and the first 150 men and 150 women in the Fall trimester selected as Sample A. The remaining 86 men and 94 women were combined with the first 64 men and 56 women from the Winter trimester to constitute Sample B. The next 100 men from the Winter trimester plus 58 women from the Winter and 42 women from the Spring trimester were combined as Sample C. Thus Samples A, B, and C contained 300, 300, and 200 *S*s, evenly divided into men and women.

---

correlated with the total E and N scale scores and nine items were retained which correlated significantly with E and not with N. These items were mixed with eleven new Si items and administered to a new sample of 145 college *S*s. Again each item was correlated with total E and N scores. Seven of the nine items common to both samples (nos. 99, 229, 304, 377, 449, 450, and 547) again met the criterion in the second sample, while two items (nos. 180 and 309) did not. Four of the new eleven items also met the correlational criterion in the second sample (nos. 57, 208, 391, and 451). Nineteen items were eliminated in this screening process, either because they failed to correlate significantly with the E score or because they showed a significant correlation with the N score in at least one of the two samples (nos. 82, 111, 143, 147, 171, 201, 254, 267, 292, 309, 415, 436, 440, 446, 455, 473, 505, 521, and 564). The eleven items surviving both analyses (nos. 57, 99, 208, 229, 304, 377, 391, 449, 450, 451, and 547) were included in the present study along with nine new Si scale items not previously screened (nos. 32, 91, 126, 316, 336, 342, 357, 400, and 497). The 20-item scales showed reliabilities (Kuder-Richardson Formula 20) of .72 and .79 in the two samples of *S*s and they correlated .63 and .74 with the MPI Extraversion scale, —.18 and —.32 with the MPI Neuroticism scale, and —.24 and —.38 with Taylor's MAS. None of the *S*s used in the above described item analyses are included in the 800 *S*s of the present study.

3. *Analysis*

The 118 items of the MPI, MAS, and MMPI Si scales were intercorrelated (phi coefficients) using the responses of the *S*s of Sample A. Two factors were extracted from this correlation matrix by the principal axis method with unities in the matrix diagonals and the factors were rotated to orthogonal simple structure using Kaiser's normalized varimax criterion. The squared factor loadings of each item were multiplied by the variance of the item to derive absolute measures of the amount of factor variance contributed by the item. The 25 items with the largest variances attributable to the "social extraversion-introversion" (SEI) factor and with the smallest variances on the "emotionality" (Em) factor were selected for further analysis along with the 25 items with the largest variances on Em and smallest variances on SEI. This procedure serves the double function of selecting items that are highly loaded on only one factor and also those whose percentage of endorsement by the *S*s is closer to the desirable 50 per cent level. Items with high factor loadings, but with very low or very high endorsement percentages, contribute relatively little to the variance of the factor scores and can profitably be eliminated.

The 50 items selected above were combined with the 40 GZTS R and T items and a similar analysis was performed using the responses of the *S*s of Sample B. The 90 items were intercorrelated, two factors were extracted from the matrix, the factor loadings were rotated to simple structure, and the squared factor loadings of the items multiplied by the item variances. Using the same criterion, i.e., items with the largest variance contribution on one factor and simultaneously the smallest variance contribution to the other factor, 30 SEI and 30 Em items were selected to comprise the final form of these scales.[2]

The answer sheets of the Sample C *S*s were scored for both 30-item scales and the scores were intercorrelated for the total group and also for each sex group separately. The reliabilities (internal consistencies) of the scores were estimated using Kuder-Richardson Formula 20.

C. RESULTS

In the analysis of the data from Sample A, the first two factors accounted for an average of 20.1 per cent of the item variances. Of the 25 SEI items selected by the procedure detailed above, 15 originally came from Eysenck's

---

[2] The statistical analysis was accomplished through the facilities of the University of Pittsburgh Computation and Data Processing Center and was partially supported by the National Science Foundation grant No. G-11309.

E scale and 10 from the MMPI Si scale, while 13 of the 25 Em items were from Eysenck's N scale and 12 from Taylor's MAS. These results indicate that Eysenck's MPI scales provided more independent items for the measurement of factors SEI and Em than did the MMPI and Taylor scales since 58 per cent of the Eysenck items survived this first screening while only 31 per cent of the items from the other two scales were retained.

The two factors extracted in the analysis of the Sample B data accounted for an average of 18.6 per cent of the item variances. All 25 of the SEI items selected in the first analysis also survived the second analysis with five Guilford R scale items being added to comprise the final 30-item SEI scale. Two of the 25 Em items (one Eysenck and one Taylor item) were eliminated in the second analysis, so that the final 30-item Em scale contained 12 Eysenck N scale items, 11 Taylor MAS items, and seven Guilford T scale items. Parenthetically it might be noted that the two factors obtained in the Sample B analysis were also rotated to oblique simple structure using the oblimax criterion of Pinzka and Saunders and the correlation between the two best fitting oblique factors was only —.10, indicating that the use of the orthogonal normalized varimax rotation criterion gave results similar to an oblique solution and that the SEI and Em factors are essentially independent or orthogonal as has been claimed by Eysenck. The item content of the derived scales can be found in Table 1.[3] In parentheses after each item is given the source (E for Eysenck's MPI, T for Taylor's MAS, M for the MMPI Si scale, and G for the GZTS) and item number (the MPI for E items, the GZTS for G items, and the MMPI booklet for M and T items). All items are scored for "true" responses except those marked with an asterisk.

The answer sheets of the 200 *S*s in Sample C were scored for the 30-item SEI and Em scales and the SEI scores were found to range from 2 to 30 for both men and women *S*s while the Em scores ranged from 0 to 29 within both sex groups. The mean SEI scores for men and women (17.60 and 17.21) were not significantly different ($t = .42$), nor were the means (14.16 and 14.60) of the Em scale ($t = .44$). The standard deviations of the SEI (6.49 and 6.51) and Em (7.07 and 6.92) scales were also not significantly different ($F = 1.01$ and 1.05). The Kuder-Richardson Formula 20 reliability coefficients of the SEI and Em scales were .87 and .89 for men, .88 and .88 for women, and .88 and .88 for the total group. The intercorre-

---

[3] Complete tables giving the rotated factor loadings of the items included in the Sample A and Sample B analyses are available from the author upon request.

TABLE 1

ITEM CONTENT OF THE PITTSBURGH SOCIAL EXTRAVERSION-INTROVERSION (SEI) AND EMOTIONALITY (Em) SCALES

Social Extraversion-Introversion Scale

1. I am happiest when I get involved in some project that calls for rapid action. (E-1)
2. I usually take the initiative in making new friends. (E-4)
3. I would rate myself as a lively individual. (E-8)
4. I would be very unhappy if I were prevented from making numerous social contacts. (E-9)
5.* I am inclined to keep in the background on social occasions. (E-14)
6. I like to mix socially with people. (E-20)
7.* I am inclined to limit my acquaintances to a select few. (E-22)
8. I like to have many social engagements. (E-26)
9. I generally prefer to take the lead in group activities. (E-28)
10. I nearly always have a "ready answer" for remarks directed at me. (E-32)
11. I would rate myself as a happy-go-lucky individual. (E-34)
12.* I am inclined to keep quiet when out in a social group. (E-36)
13. I can usually let myself go and have a hilariously good time at a gay party. (E-38)
14. Other people regard me as a lively individual. (E-44)
15. I would rate myself as a talkative individual. (E-46)
16. I am a good mixer. (M-57)
17. I like to go to parties and other affairs where there is lots of loud fun. (M-99)
18. I like to flirt. (M-208)
19. I should like to belong to several clubs or lodges. (M-229)
20.* At parties I am more likely to sit by myself or with just one other person than to join in with the crowd. (M-377)
21. I love to go to dances. (M-391)
22. I enjoy social gatherings just to be with people. (M-449)
23. I enjoy the excitement of a crowd. (M-450)
24. My worries seem to disappear when I get into a crowd of lively friends. (M-451)
25. I like parties and socials. (M-547)
26. I am a carefree individual. (G-32)
27. I make decisions on the spur of the moment. (G-67)
28. I like wild enthusiasm, sometimes to a point bordering on rowdyism, at a football or baseball game. (G-117)
29. I generally feel as though I haven't a care in the world. (G-132)
30. I usually say what I feel like saying at the moment. (G-137)

Emotionality Scale

1. I am sometimes happy, sometimes depressed, without any apparent reason. (E-2)
2. I am frequently "lost in thought" even when supposed to be taking part in a conversation. (E-6)
3. I am inclined to be moody. (E-10)
4. I have frequent ups and downs in mood, either with or without apparent cause. (E-11)
5. I am inclined to ponder over my past. (E-15)
6. I sometimes feel "just miserable" for no good reason at all. (E-17)
7. I often find that I have made up my mind too late. (E-19)
8. I would rate myself as a tense or "high strung" individual. (E-27)
9. I often experience periods of loneliness. (E-29)
10. I like to indulge in a reverie (daydreaming). (E-31)

TABLE 1 (*continued*)

11. I have often felt listless and tired for no good reason. (E-35)
12. I have sometimes been bothered by having a useless thought come into my mind repeatedly. (E-41)
13. I work under a great deal of tension. (T-13)
14. I find it hard to keep my mind on a task or job. (T-32)
15. I certainly feel useless at times. (T-142)
16. I frequently find myself worrying about something. (T-217)
17. I have periods of such great restlessness that I cannot sit long in a chair. (T-238)
18. I am more sensitive than most other people. (T-317)
19. I have sometimes felt that difficulties were piling up so high that I could not overcome them. (T-397)
20.* I am usually calm and not easily upset. (T-407)
21. At times I think I am no good at all. (T-418)
22. I worry quite a bit over possible misfortunes. (T-431)
23. I sometimes feel that I am about to go to pieces. (T-555)
24. I sometimes find myself "crossing bridges before I come to them." (G-107)
25. I am philosophically inclined, that is, inclined to philosophize about things. (G-188)
26. I often take time out just to meditate about things. (G-218)
27. I try to sense what people are thinking about as they talk to me. (G-238)
28. I frequently find myself in a meditative state. (G-268)
29. I often watch others to see what effects my words or actions have upon them. (G-273)
30. I am inclined to be introspective, that, is, to analyze myself. (G-278)

* Items scored for false response.

lations between the two scales were —.12 for men, —.17 for women and —.14 for the combined sex groups.

D. Discussion

The SEI and Em scales developed in this study appear to be more adequate measures of these two second-order personality dimensions than have previously been available for research, particularly when the subjects used in experimental studies are American college students. The item content should be somewhat broader in scope than the previous scales since the items were selected from a large pool of items drawn from six scales previously developed by different experimenters from diverse sources. Each scale contains items from three previous inventories: the MPI, GZTS, and the MMPI. It is instructive to note that the "somatic" items that are so characteristic of Taylor's MAS, i.e., items referring to physical symptoms such as nausea, sweating, headaches, etc., were eliminated in the objective item selection procedure, indicating that these items either had low loadings on the Em factor or had small variances. Inspection of the data for these items indicated that both of these item characteristics were responsible for their elimination.

The SEI scale in particular seems to be an improvement over previous scales such as the MMPI Si and MPI E scales. The much longer 70-item Si scale and the slightly shorter 24-item E scale have shown reported reliabilities of .78 for Si (McQuary & Truax, 1952) and of .73 to .82 (Bendig, 1959a) when administered to American college *S*s compared to the SEI reliability of .88 found with Sample C. The reliability of the 30-item Em scale (.88) is about the same as Taylor's 50-item MAS and probably slightly higher than the 24-item MPI N scale (Bendig, 1959a). The small negative correlation (—.14) between the SEI and Em scales is slightly, but certainly not significantly, lower than the correlation between the MPI E and N scales (—.19) (Bendig, 1959a). The Em scale, like the MPI N scale, appears to involve much less of a contaminating "introversion" component than does Taylor's MAS and similar "anxiety" and "neuroticism" questionnaires (Bendig, 1960a).

It should be noted that these Pittsburgh SEI and Em scales are the only measures of the second-order "extraversion" and "emotionality" factors, with the exception of scales developed by Cattell, that were constructed by selecting items by factor analytic procedures. All other similar scales consist of items selected by correlating items with an internal criterion, such as scores of items (Eysenck's MPI), or an external criterion, such as clinical diagnosis (Winne's Neuroticism scale), or by selecting items that rationally appear to measure a defined trait (Taylor's MAS). Only Eysenck's MPI scales were factorily validated by a factor analysis and this was done only after the items had already been selected for the scales. Cattell's "extraversion" scale requires administration of his entire long 16PF inventory to the *S*s and requires much more elaborate and costly scoring procedures. Consequently, the Pittsburgh scales appear to have unique advantages for large-scale experimental studies of personality.

E. SUMMARY

A total of 158 personality items drawn from Eysenck's Maudsley Personality Inventory, Taylor's Manifest Anxiety Scale, the MMPI Si scale, and the Restraint and Thoughtfulness scales of the Guilford Zimmerman Temperament Survey were converted to a common first person declarative statement form and administered in a questionnaire to college *S*s enrolled in introductory psychology. The true-false responses to the 118 MPI, MAS and MMPI items of 300 *S*s (150 men and 150 women) were intercorrelated (phi), two principal components factors were extracted from the matrix and rotated to simple structure (normalized varimax), and the 25

best "social extraversion-introversion" and 25 best "emotionality" items selected. The responses to these 50 items, plus 40 GZTS items, were intercorrelated for a new sample of 300 *S*s (150 men and 150 women), two factors were extracted and rotated to simple structure, and the 30 best items on each factor were selected to form the final SEI and Em scales.

The item responses of a third sample of 200 *S*s (100 men and 100 women) were scored for the 30-item SEI and Em scale to determine their psychometric adequacy. No significant sex differences in means, standard deviations, reliabilities, or interscale correlation were found. For the total group (N = 200) the reliabilities of each scale were .88 and the intercorrelation was —.14. It was suggested that these scales provide more reliable, factorily valid, and independent measures of these two important second-order personality dimensions than do any already existing scales and are to be preferred for personality research.

REFERENCES

1. BENDIG, A. W. Extraversion, neuroticism, and manifest anxiety. *J. Consult. Psychol.,* 1957, **21**, 398.
2. ———. Identification of item factor patterns within the Manifest Anxiety Scale. *J. Consult. Psychol.,* 1958, **22**, 158.
3. ———. College norms for and concurrent validity of the Pittsburgh revisions of the Maudsley Personality Inventory. *J. Psychol. Stud.,* 1959, **11**, 30-35. (a)
4. ———. Score reliability of dichotomous and trichotomous item responses on the Maudsley Personality Inventory. *J. Consult. Psychol.,* 1959, **23**, 181. (b)
5. ———. Factor analyses of "anxiety" and "neuroticism" questionnaires. *J. Consult. Psychol.,* 1960, **24**, 161-168. (a)
6. ———. Item factor analyses of the scales of the Maudsley Personality Inventory. *J. Psychol. Stud.,* 1960, **11**, 104-107. (b)
7. ———. Improving the factorial purity of Guilford's Restraint and Thoughtfulness scales. *J. Consult. Psychol.,* 1961 (in press).
8. ———. Factor analyses of the Guilford Zimmerman Temperament Survey and the Maudsley Personality Inventory. *J. Gen. Psychol.* (in press). (a)
9. ———. The factorial validity of the Guilford Zimmerman Temperament Survey. *J. Gen. Psychol.* (in press). (b)
10. DRAKE, L. E. A social I.E. scale for the MMPI. *J. Appl. Psychol.,* 1946, **30**, 51-54.
11. EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Rivista di Psicologia,* 1956, **50**, 113-140.
12. ———. The Maudsley Personality Inventory. London: Univ. London Press, 1959.
13. GUILFORD, J. P., & ZIMMERMAN, W. S. The Guilford Zimmerman Temperament Survey, Beverly Hills, Calif.: Sheridan Supply, 1949.
14. JENSEN, A. R. The Maudsley Personality Inventory. *Acta Psychol.,* 1958, **14**, 314-325.
15. MCQUARY, J. P., & TRUAX, W. E. A comparison of the group and individual forms of the MMPI. *J. Educ. Res.,* 1952, **45**, 609-614.

16. Taylor, J. A. A personality scale of manifest anxiety. *J. Abnorm. & Soc. Psychol.,* 1953, **48**, 285-290.

Department of Psychology
University of Pittsburgh
Pittsburgh 13, Pennsylvania

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 211-218.

VIGILANCE FOR CUTANEOUS AND AUDITORY STIMULI AS A FUNCTION OF INTERSIGNAL INTERVAL AND SIGNAL STRENGTH*[1]

Psychology Division, U.S. Army Medical Research Laboratory

GLENN R. HAWKES AND MICHEL LOEB

A. INTRODUCTION

Previous research (7, 9) has indicated that efficiency of responding to weak auditory signals over time is quite good, but responding to weak electrical cutaneous signals undergoes a progressive decrement. There are advantages, however, to monitoring cutaneous rather than visual or auditory signals, as the visual and auditory channels are frequently pre-empted by task requirements. Loeb and Schmidt (11, 12) have shown that for auditory signals there is a lowered probability of obtaining a performance decrement if signal strength is increased. It is of interest to determine whether or not a performance decrement with time on task would be obtained with cutaneous signals if a signal of moderate strength as opposed to the very weak signals of the previous studies were used. Signals of moderate as well as weak strength were used in the present study, therefore, in an attempt to specify the conditions under which cutaneous signaling would be efficient.

As indicated in a recent review of the vigilance literature (13), the evidence on influence of intersignal interval is confusing. Beinert and Graham (2) on a visual task, and Kappauf and Powe (9) on an audiovisual task, found a lowered probability of signal detection with increase in the intersignal interval. On the other hand, two recent experiments by Adams and Boulter (1) in a visual task, and by Harabedian, McGrath, and Buckner (6) with visual and auditory stimuli, found the opposite effect or no effect at all.

A lowered efficiency of responding to pure tones (in terms of response speed) as a function of time on task was found in the Loeb and Schmidt experiments cited above. A performance decrement for acoustic white noise

---

* Received in the Editorial Office on October 3, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This research was aided by a contract between the Office of The Army Surgeon General and the University of Louisville. The authors are indebted to Dr. Ray Bixler for administrative assistance and to D. Kaufman, C. Gettys, C. Roulston, and E. A. Schmidt for technical assistance.

stimuli was not found in the Hawkes and Loeb and Loeb and Hawkes studies, however, presumably due to the difference in spectrum of the stimuli. A performance decrement for response to the presence of broad-band auditory stimuli might be obtained, nonetheless, if intersignal interval was increased.

In the present study, both moderate and weak intensity signals were used for cutaneous and auditory stimuli. Intersignal interval was increased by a factor of five from those of the previous studies (7). Both arousal and filter theories (10) would predict improved performance with increased signal strength, but if intersignal interval affects vigilance then such results would be best explained on an arousal basis.

B. METHOD

Unless otherwise specified, the apparatus and procedure were identical to those of previous studies (7, 10). Auditory stimuli were produced by feeding the output of a random-noise generator through an attenuator into Permoflux PDR-8 earphones. The electrical cutaneous stimuli consisted of random-noise generator output fed through an amplifier into a 12 mm. diameter electrode taped to the index finger pad; a 25 mm. diameter inactive electrode was taped to the palm of the hand.

Each of 24 *O*s (12 male and 12 female university students) participated in two sessions involving electrical cutaneous signals and two sessions with auditory signals. None of the *O*s had previously participated in this type of experimentation. Suprathreshold signal duration was 0.5 of a second, with an onset or offset time of 50 msec. In one session for the cutaneous modality signal intensity was 1.2 db *SL* (*re:* absolute threshold) and 5.1 db *SL* in the remaining session. In sessions utilizing auditory signals, a matching procedure was employed to establish a signal strength subjectively equal to the electrical cutaneous stimuli.

The *O*s were instructed to press a telegraph key as soon as possible when the presence of a signal was detected. The interstimulus interval was 150, 225, 300, 375, 450, 525, or 600 seconds. The average intersignal interval was five times that in an earlier experiment by the present authors (7). There were two succesive blocks of seven trials each, with each of the intervals used once (in a random order) in each block. Duration of a session was approximately 100 minutes, as in the earlier experiment (10).

Earphones and electrodes were worn by all *O*s during each session. Earphones were relatively comfortable and were mounted in Willson Sound Barrier earmuffs, providing 15-25 db attenuation of ambient noise (dependent on frequency). Ambient noise level in the experimental chamber was about

25 db. Actions of *O* were monitored with the aid of a closed-circuit TV system, and voice communication between *O* and *E* was possible. The *O*s were paid a standard amount for their participation, and in addition a cash bonus was offered for the best over-all performance over the four experimental sessions.

Reaction time to the nearest msec. was recorded from a Standard Electric timer. Failure to respond to a signal within 3.0 seconds after its initiation was recorded as an error of omission. All responses made before or after such 3.0 second periods were scored as false responses.

C. RESULTS

The mean response times for each of seven blocks of two trials are plotted in Figure 1. Errors of omision were negligible (less than 1 per cent) except for the Faint Electrical Cutaneous (EF) condition. The per cent of signals missed in this condition is plotted in Figure 2 on the line labelled Slow Rate. The per cent of signals missed on corresponding blocks (of 10 trials each) in the earlier (10) experiment are plotted for comparison purposes in the same figure along the line labelled Fast Rate. The rate of presentation in the earlier experiment was five times as fast as that in the present one.

It is apparent from Figure 1 that latency for auditory signals is less

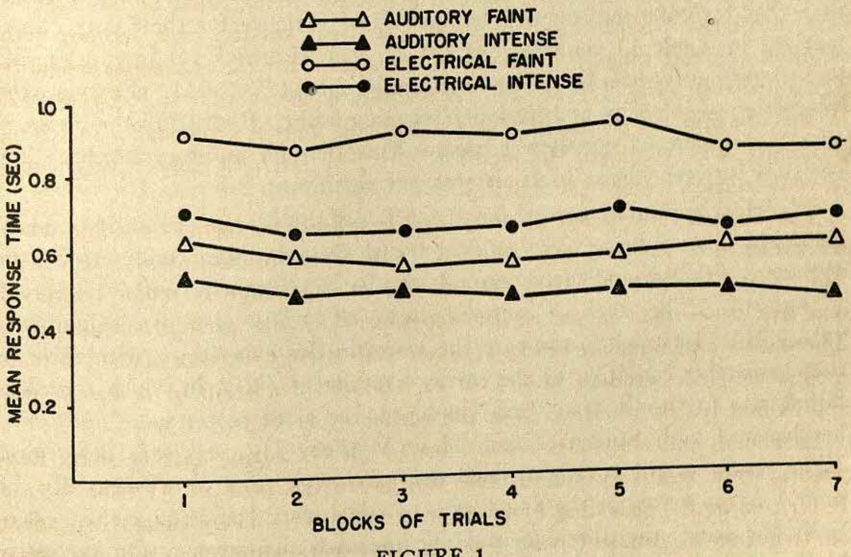


FIGURE 1
RESPONSE LATENCIES FOR THE FOUR KINDS OF SIGNALS IN EACH 90-MINUTE SESSION

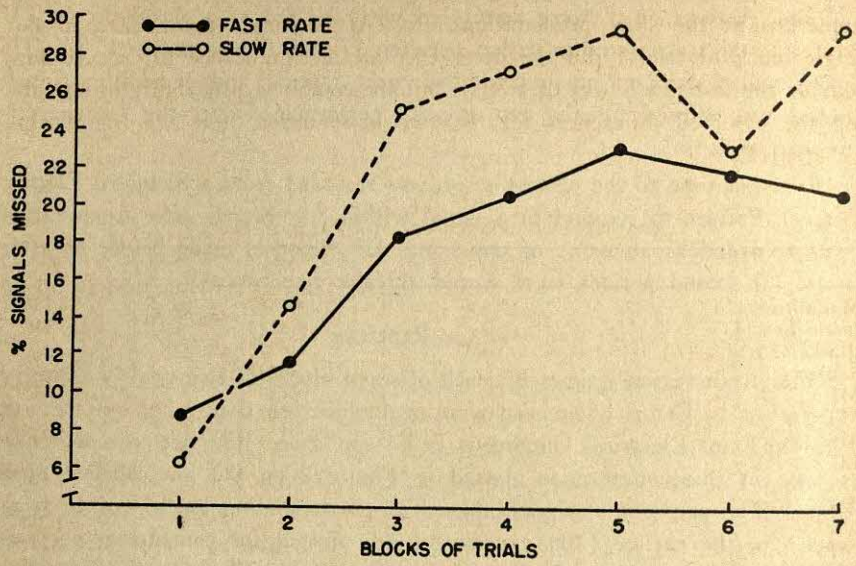


FIGURE 2
PROPORTIONS OF FAINT CUTANEOUS SIGNALS MISSED AT THE FAST PRESENTATION RATE (EARLIER EXPERIMENT) AND THE SLOW PRESENTATION RATE (PRESENT EXPERIMENT)

than that for cutaneous ones. The analysis of variance for these data (summarized in Table 1) confirms this ($P < .01$). Latency for intense signals was significantly less than that for weak ones ($P < .01$). Further, the interaction of intensity and modality was significant ($P < .01$); the intensity effect for cutaneous signals was greater than that for auditory stimuli. The variance between blocks of trials was not significant.

The distribution of errors for the EF condition was obviously skewed. Errors in this condition were compared with those in other conditions by the Wilcoxon T-test and were found to be significantly more numerous ($P < .01$). The increase across blocks of trials also was significant ($P < .01$). In order to compare the scores in this experiment with those in the comparable condition of the earlier experiment (10), in which a presentation rate five times faster was employed, the error scores were multiplied by five and both Student t and Mann-Whitney U tests were performed. (The results of the parametric and non-parametric tests were essentially the same.) Mean errors in the first blocks of trials were (percentagewise) essentially the same; the difference did not approach significance. On successive blocks of trials, the curves progressively diverged; per cent errors were more numerous at the slow presentation rate. At no single point (i.e., in no single block of trials) did the divergence attain significance ($P < .10$ for four of the last five blocks of trials), but the combined probability of obtaining the observed differences for the two presentation rates was significant ($P < .01$).

TABLE 1
ANALYSIS OF VARIANCE FOR THE LATENCY DATA

	df	MS	F
Modalities (M)	1	4.4387	32.15**
Intensities (I)	1	9.5112	55.61**
Blocks of trails (T)	6	.02707	—
Subjects	23	.3881	—
MI	1	.6308	7.83**
MT	6	.00254	—
MS	23	.13806	1.73*
IT	6	.01496	—
IS	23	.17070	2.12**
TS	138	.01448	—
MIT	6	.01408	1.26*
MIS	23	.08065	7.21**
MTS	138	.01457	1.30*
ITS	138	.01301	—
MITS	138	.01123	—

* $P < .05$
** $P < .01$

D. DISCUSSION

The most important finding from a practical viewpoint is that *O*s can maintain a high detection rate for simple electrical cutaneous signals presented at a moderately high intensity level (avoiding pain). This conclusion can be drawn only for the case where observers are to detect the presence of a cutaneous signal; it cannot be applied without reservation on the basis of the present data to the case wherein *O*s are to detect *changes* in the signals. It cannot be assumed, moreover, that if the signals are to be detected against a background of cutaneous noise (provided, e.g., by necessary contacts with solid features of the environment, air currents, etc.) that detection would be as good as in the present study or as well maintained. Detection of moderately intense cutaneous signals against a background of masking noise should be investigated. Unless evidence is presented to the contrary, however, it may be assumed that moderately intense electrical cutaneous stimuli can be reliably monitored.

The finding that wider spacing of signals (used in the present study) apparently increases the performance decrement, i.e., a decrease in detection of weak signals, is of considerable theoretical interest, though the difference in detections throughout the sessions reported herein was probably not of great practical import. It was mentioned earlier that findings regarding the effect of variations in signal presentation rate were quite variable. Some writers have reported that closer spacing along the temporal dimension enhances performance; others indicate that it has no effect, and still others that the faster rate is effective only insofar as the greater number of signals arouses the observer (1). Colquhoun's experiment (3), on the other hand, indicates that the ratio of "wanted" (relevant) signals to "unwanted" (irrelevant) signals determines the course of monitoring performance, not the absolute number of signals. (Presumably situations involving a higher ratio of wanted to unwanted signals are more arousing but this is not discussed at length.) In his studies, the wanted signals differed from the unwanted ones only in the paleness of the background on which the visual signals were presented.

In an earlier paper (10), it was suggested that observers typically have "spontaneous" cutaneous sensations somewhat resembling the low-intensity cutaneous signals. It was further hypothesized that there is habituation to these spontaneous sensations over the course of a session and that this has a detrimental effect on the perception of the faint cutaneous signals. If we provisionally accept this hypothesis—and Colquhoun's suggestion on the effects of spacing of signals—and if we view such spontaneous sensations as unwanted signals and the true signals as wanted signals, we would predict that wider spacing (along the temporal dimension) of cutaneous signals should increase the decrement observed with time on task.

An alternative explanation could be advanced in terms of Deese's expectancy theory (4). It might be suggested that the fast signal rate (almost one signal/minute) is more in line with the subjects' expectations than the slow rate (approximately one every six and one-half minutes), and verbal reactions to the task indicated that this was the case. There have been experiments (5) which indicate that greater decrements result when subjects' expectations of signal schedule are not in line with reality. The theory, however, would not by itself explain the progressive decrement observed for faint cutaneous signals in the present study. It should be noted that Deese (4) also proposed an explanation of progressive decrement in terms of the effect of restricted sensory input on the waking center, but this also, would not explain the observed differences in attending to auditory and cutaneous signals. If as Holland suggested (8), signal detections are viewed as reinforcements, then the superior performance of the fast signal schedule might be explained on the basis of the greater number of reinforcements, but this too fails to explain the other observed effects.

E. SUMMARY

Vigilance for auditory or electrical cutaneous stimuli at both weak and moderate intensities was investigated. Intersignal interval ranged from 150 to 600 seconds.

Latency for detection of auditory signals was shorter than for cutaneous ones, and shorter for signals of moderate as opposed to weak intensity. Latency, however, did not significantly increase as a function of time on task. Detection probability was very high for moderate cutaneous and auditory signals and for weak auditory signals, but it was appreciably lower for faint cutaneous signals. A significant progressive increase in failures of detection of faint cutaneous signals was found with time on task. Increase of the intersignal interval by a factor of five over that employed in a previous experiment resulted in a significant increase in the number of errors.

Possible theoretical implications of the findings are discussed.

REFERENCES

1. ADAMS, J. A., & BOULTER, L. R. Monitoring of complex visual display. I. Effects of response complexity and intersignal interval on vigilant behavior when visual load is moderate. *AF Command & Control Development Div.*, Tech. Note AFCCDD-TN-60, 63, Aviation Psychol. Lab., Univ. Illinois, 1960.
2. BARTLETT, S. G., BEINERT, R. L., & GRAHAM, J. R. Studies of visual fatigue and efficiency in radar observation. *Hobart & William Smith College Rep.*, 1955.
3. COLQUHOUN, W. P. The effect of "unwanted" signals on performance in a vigilance task. *Ergonomics*, 1961, **4**, 41-51.
4. DEESE, J. Some problems in the theory of vigilance. *Psychol. Rev.*, 1955, **62**, 359-368.
5. FLOYD, A., GRIGGS, G. D., & BAKER, R. A. The role of expectancy in auditory vigilance. U.S. Army Armor Human Res. Unit Rep. (in press).
6. HARABEDIAN, A., McGRATH, J. J., & BUCKNER, D. N. The probability of signal detection in a vigilance task as a function of intersignal interval. *Human Factors Res., Inc. Tech. Rep. 3*, Los Angeles, Calif., 1960.
7. HAWKES, G. R., & LOEB, M. Vigilance for cutaneous and auditory signals. *J. Aud. Res.*, 1961, **1**, 272-284.
8. HOLLAND, J. G. Human vigilance. *Science*, 1958, **128**, 61-67.
9. KAPPAUF, W. E., & POWE, W. E. Performance decrement on an audio-visual checking task. *J. Exp. Psychol.*, 1959, **57**, 49-56.
10. LOEB, M., & HAWKES, G. R. The effect of rise and decay time on vigilance for weak auditory and cutaneous stimuli. *Percept. & Mot. Skills* (in press).
11. LOEB, M., & SCHMIDT, E. A. Influence of time on task and false information on efficiency of responding to pure tones. *USA Med. Res. Lab. Rep.* No. 426, Ft. Knox, Ky., 1960.

12. ———. A comparison of the effects of different kinds of information in maintaining efficiency on an auditory monitoring task. *USA Med. Res. Lab. Rep.* No. 453, Ft. Knox, Ky., 1960.

13. McGrath, J. J., Harabedian, A., & Buckner, D. N. Review and critique of the literature on vigilance performance. *Human Factors Res., Inc. Tech. Rep.* 1, Los Angeles, Calif., 1959.

Sensory Psychology Section
Civil Aeromedical Research Institute
Federal Aviation Agency
Oklahoma City, Oklahoma

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 219-232.

THE USES AND INTERPRETATIONS OF A FOLLOW-UP STUDY OF PATIENTS IN PSYCHOANALYTIC PSYCHOTHERAPY*

Postgraduate Center for Psychotherapy and Department of Psychiatry, New York Medical College

BERNARD F. RIESS, RALPH GUNDLACH, AND CLIFFORD J. SAGER

A. PURPOSE

Researchers in psychotherapy have traditionally differentiated between process and outcome studies. The former have had much greater status and prestige because they are allegedly more basic, more purely scientific and closer to important problems of personality structure and dynamics. Outcome studies are denigrated as bread-and-butter investigations, special pleadings for a given mode of treatment and experimentally naive. It is true that the reports on outcome have run a tremendous gamut from single-case samples to ambitious projects involving a plethora of major and minor hypotheses, experimental strategies and logistics resembling those employed in the construction of an outer space vehicle. More and more, the investigation of the effect of a given treatment is enveloped in the need for elaborate research design and large budgets so that the unsupported, individual researchers find it impossible to conceive of engaging in outcome study. Meanwhile the therapists find themselves embroiled in the controversy between the Eysencks and the anti-Eysencks without much evidence on either side to serve as intellectual ammunition.

Recently a book by Reznikoff and Toomey (2) and a thoughtful paper by Zax and Klein (3) have summarized many of the methodological issues in the evaluation of changes associated with treatment. The general conclusion from their work is that no study with adequate control over even a majority of the already defined variables has yet been done or is apt to be done in the near future.

In view of these warnings, it may seem presumptuous to offer still another study of outcome. This study is, however, different. It deals not so much with the determination of the relationship between certain types of treatment and certain kinds of changes but rather with a fact-gathering attempt to

---

* Received in the Editorial Office on October 13, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

find out what patients were like some years after the termination of treatment. It differs from other studies in that all patients who terminated treatment in a given period were followed regardless of the success or failure of the previous therapy. Furthermore, it differs from most other studies in that the patients involved in it had had extensive periods of therapy rather than the short-term treatment often found in university or hospital clinics. Finally, the patients studied had all been exposed to psychoanalytically oriented methods as practiced by a closely identified group of therapists, who maintained contact with the patient over the whole course of his treatment. Part of our purpose in this paper is to show that such collecting of data when accompanied with a liberal dose of exploratory drive can compensate in the fruitfulness of its results and in the stimulation of ongoing search for elaborate design and huge research budget allocations. It is hoped that our enthusiasm for nose-counting will lead other investigators to rub noses with us in the so-called primitive custom of friendly cooperation.

It should not be inferred from the above that we are unaware or negligent of the experimental pitfalls and deficiencies which are imbedded in our study. We believe, however, that the existence of such factors and the difficulties of adequate control over them do not require suspension of investigation until they can be corrected but rather that the corrections should emerge from actual immersion in the research process. If one were to wait for the opportunity to employ a perfected research design in this area, the delay would be endless.

B. Conditions

The major specifications for outcome research are easily listed: (*a*) an adequate, reliable and valid estimate of initial status of the patient at the time of onset of treatment; (*b*) an adequate representation of the parameters of the treatment process; (*c*) an adequate estimate of the personality, orientation and attitudes of the therapist as well as his treatment procedures; (*d*) a control group of untreated, dynamically matched patients whose major problem areas are equivalent to the experimental group; (*e*) an adequate, valid and reliable estimate of the patient's status at the termination of treatment to be compared with the same estimate for the control group; and (*f*) an independent estimate of the effect of treatment or the durability of change after the passage of non-therapeutic time. Implied in the above stages is the need for clear-cut and agreed-upon criteria of change, personality status and good mental health. Finally, there must be devised some procedures for taking into account the placebo effect, the therapist's communication to the patient of his, the therapist's faith in the efficacy of what he practices.

The sources of criteria in most studies of outcome can be classified in several ways. Some are derived from the patient's reports about himself as to his intra-psychic status, feelings, symptoms and self-attitudes. Others get at the same data from the viewpoint of the treating personnel in the form of reports, interviews and questionnaires. Another cross-section gets at criteria that refer to behavior in situations outside the therapy itself such as life, family, social and work attitudes and interactions. Most of the studies in the comprehensive coverage of the two references cited above have their locale in either university clinics or in-patient hospitals where the patients and therapists are available for extensive and intensive questioning by Q-sorts, rating scales or personal depth interviews. In the private or public agency with an out-patient setting, there is little opportunity for this time-consuming procedure nor can patients be required to participate in such designs. Hence there must be recourse to less time-consuming and expensive techniques.

The setting for our study is a large, out-patient, psychoanalytically oriented treatment center. More than 10,000 patients have been seen at the Postgraduate Center for Psychotherapy during its sixteen years, and with a frequency between one and three sessions per week. Fees are set at a level which is available to lower and middle income persons. The therapists are either full-time fellows-in-training or staff personnel who give some seven hours a week to therapy and supervision. The three major disciplines are represented among the therapists, psychiatrists accounting for about 40 per cent of the total, psychologists 40 per cent and social workers 20 per cent.

C. SUBJECTS

The subjects for the present study were 201 cases randomly chosen from among those patients who had had at least one year of therapy at the Center and who had terminated their treatment at least two years prior to the first contact with the follow-up project. "Termination" was defined as the cessation of treatment at the Center regardless of whether such termination was planned or unplanned and independent of any evaluation of the patient's past progress or status at the time of termination. In other words, the subjects are simply persons with at least one year of therapeutic experience followed by at least two years of non-appearance on the Center's patient lists.

D. Method

Data about these subjects were collected at four points on their life continuum, starting with first contact with the Center and concluding with the end of the investigation. The first set of data, derived at the moment of application at the Center, consists of material on the psychiatric screening interview report and on a Personal History Sheet filled out by the patient. The screening report is an evaluation of the patient by a trained psychiatrist who is asked to record initial complaint, make diagnoses and prognoses, estimate tentative dynamics and recommend type and frequency of treatment. A brief anamnesis is also obtained by the screener. The Personal History Sheet is an elaborate questionnaire which develops a complete social, family and work history of the patient and his or her immediate family. Both records are prepared so that material can be entered on IBM cards.

The second point on the continuum coincides with the cessation of treatment. The record here is a termination report by the patient's therapist after discussion with the therapist's supervisor. It consists of an evaluation of present status, change if any from the initial condition of the patient, prognosis and a brief description of the course of therapy including the final psychodynamic picture of the patient as seen by the therapist.

The third point was a questionnaire sent to patients after at least two years of separation from the Center. This document, forwarded under the signature of the therapist, was designed to probe the following areas: (*a*) symptomatology as recalled at the onset of treatment and at present; (*b*) current sense of being; (*c*) relationships with a variety of significant fellow-individuals, mates, co-workers, superiors, subordinates, friends, parents and children; (*d*) ability to function independently at work, socially and sexually; (*e*) attitudes towards therapeutic experiences; and (*f*) projections about self that would allow for evaluations by the research staff.

The final point was a follow-up interview administered to a selected group of respondents from among the 201 by a psychiatrist with no previous contact with either the patients or the research.

In summary then, the patient is scrutinized at three points in his life, onset of therapeutic experience, at the end of the experience and at the present, some two years or more after termination. In this scrutiny, evaluations are made by five different and independent observers; the patient, the screening psychiatrist, the therapist, the research staff and the follow-up psychiatrist.

There are several assumptions underlying our methodology. In the first place, we assume that there is no generally agreed upon instrument with which validly to assess the initial status of the patient. For this reason we rely heavily on the patient's self-report of what he said was the reason for seeking help initially. The recall some three years later of the initiating complaint should offer an intriguing comparison not only of the effect of treatment but of the increased sophistication and/or self-understanding of the patient. Second, we assume that the effect of treatment should be visible in the behavioral adjustments of the patient. Hence the emphasis on the socio-economic and psycho-social aspects of the patient's reports, both pre- and post-treatment. Another assumption is made which relates to the validity of psychoanalytic training, namely, that trained analysts should be able to abstract personality changes from the reports of the patients about themselves and that agreement as to these changes should be possible among trained observers. Finally, we assume that the nature of the therapeutic relationship obtaining at the Center would facilitate the obtaining of a representative sample of the population of terminated cases.

E. RESULTS

1. *Questionnaire Returns*

Let us now turn to the actual results of the study. First, we have demonstrated that, under the conditions of relationships at the Center, a mail questionnaire will yield substantial returns. This is shown in Table 1 in which the ultimate fate of the 201 mailed questionnaires is described. It will be noted that of the 201, 108 (53 per cent) were returned. Of these, five were merely signed and sent back with no answers. Of the remainder, 62 questionnaires did not reach their subjects because of change of address

TABLE 1
FREQUENCY OF RETURN OF QUESTIONNAIRES BY VARIOUS GROUPS

	N	% of 201	% of 135
Total population	201		
A. Returned with answers (Group R)	103	51	76
Returned blank	5	2	3
B. Returned incorrect address (Group IA)	62	31	—
C. Not returned (Group NR)	27	13	20
D. Not returned, but information given by other sources	4	2	

and removal without notification of new address. Thus, only 31 individuals actually received the document and refused to return it. In this group we subsequently learned that there were two deceased former patients, an additional one whose current therapist encouraged the patient not to respond and one who would have responded but was unreachable. In summary, 76 per cent of those who received the questionnaire reacted to it in a way so as to afford data for research with an additional 3 per cent who reacted but gave little or no information.

This relatively high frequency of returns raises methodological queries. One is whether the total group of 201 was really representative of the patient population or selected in some systematic fashion despite the apparent randomization of choice. An answer to this can be made by comparing the 201 cases with the over-all patient population of the Center. The latter, numbering 657 patients who were in treatment in 1957, was found to differ significantly in no single item from the 201 terminees. Comparison items were the data on the Initial Interview and the Personal History Sheet. Contamination was avoided since the 201 experimental subjects were not included in the control group of 657. Thus, it is fair to say that the population who were studied by us were representative, at the time of entrance into treatment, of the general population at the Center.

The second methodological problem is whether the group which gave us usable responses, the 103 called the "R" group, was different from the groups receiving the material but not cooperating in the study, the "NR" group and those whose mobility prevented them from getting the questionnaire, the "IA" population. The relevant data are presented in Tables 2 and 3. In Table 3 are given the results both for the material from the Personal History Sheet and the length of therapeutic experience prior to termination. It will be seen that there are a few significant differences in the sociological items. Only those factors showing some differences are given in the first fourth of the table, together with the level of significance of the differences. All other sociological items were essentially the same for the three groups. It is apparent that the group with higher mobility, the IA class, was the most deviant although no consistent trend is visible.

The middle two banks of data in Table 2 deal with the duration of therapy and the total number of therapeutic sessions. Here the only significant differences are between the non-respondents and the incorrect address groups at the modal duration of treatment. Considering all three subgroups of the study population, it is clear that therapy was relatively long in duration and extent with over one third of all subjects having more than

TABLE 2
COMPARISONS AMONG THREE STUDY SUBGROUPS—IN PER CENTS

Personal history sheet data	Group R	Group NR	Group IA	Significant difference
	(N103)	(N 31)	(N 62)	
Sex: Males	40	29	60	R-IA and
Females	60	71	40	NR-IA 5%
Living with: Self	12	23	27	R-IA 5%
Spouse	31	23	15	R-IA 5%
Father's Occupation:				
Small business, or Minor executive	14	16	39	R-IA 5%
Skilled or Semi-Skilled worker	23	35	13	
Place of Birth: USA	80	65	95	NR-IA 5%
Father, USA	19	10	32	
Mother, USA	19	6	40	NR-IA, R-IA 5%
Mother, Eastern Europe	35	48	18	NR-IA, R-IA 5%
Duration of Therapy	(N108)	(N 31)	(N 62)	
1-1.5 years	28	35	32	
1.5-2.0 years	35	32	34	
2.0-2.5 years	9	3	18	NR-IA 5%
2.5-3.0 years	9	6	3	
3.0 or more years	18	23	13	
Total Number of Sessions	(N107)	(N 30)	(N 60)	
31-50	7	3	8	
51-75	17	13	12	
76-100	12	10	22	
101-150	31	40	28	NR-IA 5%
151-200	17	23	20	
201-299	9	7	8	
300-399	6	0	2	
400 or more	1	3	0	
Type of Termination	(N105)	(N 31)	(N 62)	
Planned	64	45	58	R-NR 5%
Unplanned	36	55	42	NR-IA 5%

two years of treatment and more than half having over 101 sessions. Thus we are dealing with people whose therapeutic experience is considerable. Finally, the type of termination, given in the last bank of the table, indicates that there were significantly more planned cessations of therapy for the respondent than for the other groups, particularly than for the non-respondent class. This may be a partial explanation of the resistance of this section to cooperate in the investigation.

This fact, of a difference in planned and unplanned terminations, does not however mean that the patients in the uncooperative group were felt by their therapists to be less successful in treatment. Table 3 shows the estimates by the therapists of the extent of change in the three subgroups. Over two-thirds of all groups were seen by their therapists as moderately or more than moderately improved.

TABLE 3
COMPARISONS AMONG THREE SUBGROUPS IN TERMS OF THERAPISTS' ESTIMATES OF CHANGE

Estimate of change	Group R (104) N	Group R (104) %	Group NR (30) N	Group NR (30) %	Group IA (61) N	Group IA (61) %	Significant difference
Recovered	3	3	0	—	2	3	
Marked improvement	39	38	10	33	23	38	
Moderate improvement	45	43	12	40	24	39	
Slight improvement	15	14	7	23	7	11	NR-IA 5%
Unimproved	2	3	1	3	4	7	
Worse	0	—	0	1	1	2	

In summary of the methodological questions raised above, then, it is possible to say that there are few significant differences among the groups to whom the questionnaire was sent regardless of what they did with the document. Thus we can say with a fair degree of confidence that the conclusions from the population of 103 who answered will hold for the total sample of terminated cases.

2. *Estimates of Evaluators*

We turn now to the consideration of what happened to the respondent group. We have already noted the therapist's estimate of improvement for this group in comparison with the other subject populations. However, there are two other sets of estimates for the patients in Group R. One is that by the patient himself as given on the questionnaire. The third evaluation is that of the researchers on the basis of their over-all reading of the questionnaire responses. Before making these three-fold comparisons a word should be said about the evaluations by the researchers. The procedure was as follows: Each research member individually and independently went over the whole questionnaire for each patient and estimated the amount of change in terms of a four-point scale ranging from marked improvement through unchanged to worse. After all 103 returns had been so evaluated, the researchers met together and discussed those questionnaires on which there was a two-point or greater disagreement. After discussion, each participant re-rated the questionnaire. The inter-rater reliability was considerable since two-point disagreement existed on only 12 of the 103 cases. Where there was a one-point disagreement, the patient was given the lower rating.

Table 4 gives the estimates by the therapists and researchers for the over-all impression of change whereas the patient's evaluations are given in response to questions as to the change in the presenting symptoms and as to current functioning in comparison with past behavior. Inspection of the results shows that the patient's self-estimates are remarkably consistent when asked about changes in the presenting symptoms and the level of present functioning. There are significant differences however between the patient's self-concepts and the opinions of their therapists and of the researchers. Both these evaluators are not as optimistic about marked improvement and more cautious in seeing change so that the estimates tend to bulk in the category of moderate or some improvement. However, there is general agreement that most of the patients in this group showed significant change in the direction of better functioning.

TABLE 4
ESTIMATES OF CHANGE IN GROUP R BY THERAPISTS, PATIENTS AND RESEARCHERS

	By Patients				By Therapist		By Researcher	
	Current function (N102)		Orig. symptom status (N100)		Final estim. of change (N104)		Over-all change (N103)	
Evaluation of change	N	%	N	%	N	%	N	%
Marked improvement	60	59	65	65	42	41	26	25
Some improvement	33	32	26	26	60	57	65	63
Unchange	7	7	7	7	2	3	11	11
Worse	2	2	2	2	0	—	1	1

Since the patients' self-evaluations were so consistent, it was thought to see whether there were any internal evidences of contradictions as to the improvement asserted by the patient. To test this, the patients' responses to questions about how they felt during the past year, whether they had developed new symptoms and whether they had made progress in work were studied. These are presented in Table 5. It is evident that the patients felt that they had functioned only fairly well during the past year and had made only satisfactory progress in work rather than being completely well or enthusiastic about their jobs. On the other hand, 22 per cent had developed new symptoms since leaving the Center. Thus it would seem that the position of both therapists and researchers came closer to the patients' status than did the over-all opinion of the patients themselves.

TABLE 5
GROUP R SELF-EVALUATIONS AS TO FEELINGS IN PAST YEAR, DEVELOPMENT OF NEW SYMPTOMS AND PROGRESS IN WORK

Evaluations*	1. Felt in past year (N 100)		2. New symptoms (N 93)		3. Progress in work (N 75)	
	n	%	n	%	n	%
Well (1) Yes (2) Enthusiastic (3)	35	35	20	22	2	3
Fair (1) Satisfied (3)	47	47			34	4
Ambivalent (1) Neutral (3)	5	5			13	17
Worse (1), no (2) dissatisfied (3)	13	13	73	78	17	22
Unemployed (3)					11	14

* 1, 2, 3 refer to the three sets of evaluations, i.e., 1 refers to how patients felt in past year, 2 to new symptoms, 3 progress in work.

Finally we have tabulated the patients' estimates of the changes in specific areas. Table 6 shows the areas and the numbers and percentages found in each degree of change within the area. The greatest area of positive change

TABLE 6
PATIENTS EVALUATIONS OF CHANGE IN SPECIFIC AREAS OF FUNCTIONING

Area of functioning	Area	Better		Same		Worse		No problem
	N	N	%	N	%	N	%	N
Marriage	51	36	71	10	20	5	10	4
Children	26	13	67	10	38	1	4	8
Co-Workers	69	50	72	16	23	3	4	18
Superiors	74	60	81	12	16	2	3	11
Subordinates	56	35	63	18	32	3	5	18
Mother	76	51	67	22	29	3	4	0
Father	60	38	63	18	30	4	7	3
Male Friends	83	51	66	29	35	3	4	13
Female Friends	81	54	67	22	27	5	6	18

Percentages are per cents of area N, omitting no problems.

was relationships with superiors where 81 per cent of those answering in this category reported that they did better. At the other extreme with relatively small percentages of improvement were relationships with male and female friends, subordinates, fathers, mothers and children. The only area with a relatively large change in the direction of worse was marriage. Hence there is some question as to whether worse refers to the dissolution of the marriage or to the attitude toward the spouse.

F. DISCUSSION

Enough of tables and statistics. We have dwelt on these in order to show some of the intricacies of relationships which can be evolved out of even the small number of subjects we have studied. A large-scale project would generate a life-time of digital manipulation and we must avoid such digital fetichism. What have we learned from our study and what ideas has it stimulated? One of our basic concepts at the Postgraduate Center is that research must have some feed-back into clinic operations and treatment procedures if it is to be meaningful within the specified setting. Therefore we shall emphasize those aspects of our scrutiny of and work with our data which seem to us to have significance for our therapeutic work.

The material of the study shows that psychoanalytically-oriented therapy produces changes in patients as evaluated by the patients, by their therapists, by researchers studying the patients' reports and by an independent psychiatric interviewer. It also shows that such therapy, carried on at a frequency of one or two sessions per week is not short-term therapy. When we study the relationship between patients', therapists' and researchers' estimates of improvement and the duration of therapy in years and number of sessions, the greatest frequency of improvement occurs in those patients who have had more than one and one-half years of treatment and more than 100 treatment sessions. Can we say, however, that this degree of improvement was satisfactory to the patient? Is termination, whether planned or unplanned, a synonym for successful treatment or cure? Our questionnaire asked patients whether they had had additional therapy since termination and whether they felt that they would like more therapy. Over half of the respondents had some additional therapy or were still in treatment and a large number felt that more help would be desirable. It is obvious, therefore, that the meaning of termination is not the same as cure even where the cessation of treatment is jointly discussed between patient and therapist. We have some data to indicate that there is a slight distortion in our data because some of the therapists in training ended their training simultaneously with the period of major patient terminations. We do not know how many of the terminees continued with their therapists in private practice. There are, however, exceptions to this in that many patients did not terminate when their therapists went into private practice and ended their training. For these patients, it is possible to speculate that both patient and therapist felt that the improvement in individual treatment had reached the point of low return. This implication has led us to discuss the possibility of working through with patients to the acceptance of new types of treatment such as group therapy or an even newer development that we call social therapy.

It will surprise no one, we are sure, to hear that there seems to be little if any relationship between the diagnostic categories of the American Psychiatric Association nosology and any datum on the questionnaire. It did strike us as important however to see what happened to the initial diagnosis after a long experience with the patient. We approached this in two ways, one by an examination of the diagnosis at initial screening and at termination and one by comparison of the patient's verbatim complaint at the time of screening and his recollection of why he had come for treatment as expressed on the follow-up questionnaire.

From the first set of comparisons, initial versus terminal diagnosis, it is clear that the policy of the clinic of cautious admission of schizophrenics was put into effect. Initially, we had only seven people classified as schizophrenics out of the total of 201. However on termination, this group had more than tripled in size, numbering 24. This should not be hastily viewed as an effect of therapy, that non-schizophrenics became schizophrenic. If we look at the source of these changed diagnoses, it appears that the majority came from the categories of passive-aggressive and character disorder. No more need to be said than to point to the breadbasket nature of these latter diagnoses.

Examination of other relationships with initial diagnoses indicates that the label of schizophrenia does not deter therapists from working with the group. As a matter of record, the diagnostic category showing the greatest frequency of marked improvement by all three evaluators was that of schizophrenia. However, the patient and researchers' estimates are closer together than those of the therapists. Since the former two evaluators based their ratings on a period two years after termination, the difference between therapists and patient-researcher may reflect the holding power or durability of change. The same trend is true of patients evaluated by their therapists as schizophrenics at termination. The effects of treatment for this group must have been considerable. The only group showing a trend from marked improvement to unchanged during the two-year follow-up period was the obsessive-compulsive.

The examination of the shifts in the patient's picture of why he originally came for treatment is interesting. Whereas the initiating complaint is usually expressed in strictly symptomatic terms, headaches, troubles with spouse, loneliness, inability to get along with others and the like, after termination, there is a marked increase in the frequency of use of psychoanalytic terminology. The presenting picture as recalled is expressed in words such as "could not express my hostility," "unresolved oedipal situation," "guilt," "sense of worthlessness." Thus, at the worst estimate, the analytic experience was educational. At the best, one may interpret the results as showing a real change in self-evaluation and understanding.

It has often been said that the evaluation of change by both the therapist and the patient is subject to a subtle socio-psychological influence. Hathaway (1) has called this the "hello-goodbye" effect. The patient-therapist relationship builds up into itself a conformity to social norms so that it becomes a matter of politeness and regard to let the therapist know that he has helped the patient. Conversely the therapist wants the patient to feel appreciated as "good" and so finds positive attitudes in the treatment results. These biases have made for difficulty in previous outcome studies because they are so resistant to exposure. In our study we had a double opportunity to get some suggestive data on the problem. If we examine the patient and therapist evaluations of treatment, we find that the patients saw themselves most frequently as markedly improved, the therapists somewhat less in agreement with this assessment but not significantly so. On the other hand the researchers were significantly more cautious about marked improvement and found greater evidence of slighter change. Second, when the "R" group patients were divided into those who had had no additional therapy, some additional treatment or were still in treatment, all evaluations found most improvement in patients in the first subdivision, next in those with some additional help and least among those still in treatment.

For those patients who had no additional therapy, the factor of planned or unplanned termination reveals the effect of possible bias by patient and therapist. Both sets of evaluators, patients and therapists, found that planned termination was accompanied by a high percentage of marked improvement whereas the researchers failed to find this to be so.

We have not thus far referred to the final point on our experimental continuum in the patients' lives, the follow-up interview. The number so seen totalled 25. Examination of the initial interview, personal history data, diagnostic categories and length of exposure to treatment indicates that these patients are not different from the remainder of the experimental population. The interview was structured to get at both the dynamic factors and the life-experiential data. An immediate result of the interview was that the patients generally focussed on changes in living status rather than on intra-psychic components. The over-all impression from the interviews substantiated the material from the questionnaires despite the difference in the evaluating personnel.

G. SUMMARY

In conclusion, we would emphasize that our study represents an attempt to find out if and where changes occur rather than why they take place. Second, we have studied people who have had extensive exposure to treatment. Third, we have attempted an extensive survey of a relatively large population. Our goals were clear and simple. We sought to find out whether experience in therapy was accompanied by changes in the relationships of patients to their way of life. Admittedly, our instruments in this project are fallible, but their lack of reliability cannot account for the large amount of agreement in the findings. Furthermore we believe we have shown that extensive investigation can expose data of significance to the therapist, especially in focussing attention on those areas of change which patients point to when questioned about the effects of treatment.

REFERENCES

1. HATHAWAY, S. R. Some considerations relative to nondirective psychotherapy on counselling. *J. Clin. Psychol.,* 1948, **4**, 226-231.
2. REZNIKOFF, M., & TOOMEY, L. A. Evaluation of Changes Associated with Psychiatric Treatment. Springfield, Ill.: Thomas, 1959. Pp. 2-132.
3. ZAX, M., & KLEIN, A. Measurement of personality and behavior changes following psychotherapy. *Psychol. Bull.,* 1960, **57**, 435-448.

Postgraduate Center
for Psychotherapy, Inc.
218 East 70 Street
New York 21, New York

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 233-244.

SOCIAL INFLUENCE ON THE RESPONSE TO DRUGS: I. AMPHETAMINE SULFATE*

Department of Psychology, University of South Dakota

NORMAN W. HEIMSTRA[1]

A. INTRODUCTION

Research in behavioral pharmacology at the subhuman level is typically concerned with the effects of a drug on some component of behavior of an organism in a solitary situation. There is some evidence to indicate, however, that social factors may influence an animal's response to a drug, Chance (2, 3); Gunn and Gurd (4). At the human level, Nowlis and Nowlis (11) have reported that certain effects of drugs may be altered by group conditions.

Several investigations have utilized some aspect of social behavior as a dependent variable in the assessment of drug effects on behavior, Norton and deBeer (9); Norton (8); Norton and Tamburro (10); Heimstra (5, 6). Since many drugs are used specifically to modify behavior in a social context, it would seem that systematic investigations of drug effects on animal behavior in a social situation would, at least in some instances, be of more value than behavioral data obtained from studies of isolated organisms.

The purpose of the present investigation was to determine the effects of social stimulation on the rat's response to amphetamine sulfate.[2] This drug was utilized for the initial study in this series since previous investigations (2, 3, 4) showed that social or aggregate factors may assume considerable importance in an animal's response to a CNS stimulant.

B. METHODS

1. *Subjects*

Forty-eight male albino rats were used as subjects. The animals were 70 days old at the start of the investigation and had been housed in individual cages from 45 days of age. Food and water were available on an

---

* Received in the Editorial Office on October 15, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This investigation was carried out during the tenure of a Postdoctoral Fellowship (MF-8654-C2) from the National Institute of Mental Health, U.S. Public Health Service.

[2] Supplied by Smith, Kline, and French Laboratories, Philadelphia.

ad libitum schedule. In addition to the animals used as subjects, a number of rats from the Psychology Department Rat Colony were used as stimuli in various conditions of the study.

2. *Apparatus and Procedures*

The general procedures and apparatus utilized were somewhat similar to those used by Bindra and Baran (1). Four identical observation boxes, with inside measurements of 24 × 12 × 12 inches, were constructed from plywood. The boxes were fitted with hardware cloth tops and bottoms. A hardware cloth barrier (.25 inch mesh) was placed in the center of each box dividing the box into two equally sized compartments. Each box was placed on a stand with the bottom of the box 11 inches from the floor. Illumination for observations was provided by a 300 watt incandescent bulb enclosed within the globe of a ceiling fixture located directly above the observation boxes.

A Stenograph machine was used to record the observations made of an animal during an observation period. Use of this machine for recording behavior observations has been described elsewhere (5, 6, 7). A key on the machine represented a category of behavior for the animal being observed. When the animal was engaged in that category of behavior, the key was depressed at one second intervals as marked by beats from an electric metronome. On a machine of this type, recordings are made on a paper tape that is advanced each time a key is depressed. Data recorded in this manner can be analyzed in several ways. In the present study, the frequency of recorded observations in a given category of behavior during an observation period was utilized. The number of behavior shifts from one category of behavior to another was also obtained.

During the observation periods, *O* sat with the Stenograph machine on a stand directly in front of him. The animal to be observed was placed in a compartment of the box which was located beside *O* and the observations were made through the hardware cloth top.

The 48 rats used as *S*s were randomly assigned to one of six groups of eight rats each. The *S*s in three of the groups were observed singly and *S*s in the remaining groups were observed in the presence of other rats used as stimuli. The *S*s in the former groups were designated solitary (Sol.) and those in the latter groups social (Soc.). All *S*s were observed in one compartment of the observation box. The second compartment was empty when the solitary *S*s were observed and contained three rats when the social rats were observed.

Animals in one of the solitary groups were given saline (S-Sol.), one group was injected with 0.5 mg/kg of amphetamine sulfate (0.5-Sol.), while the remaining group received 2.0 mg/kg of the drug (2.0-Sol.). Animals in the social groups received similar treatment with one group receiving saline (S-Soc.), one group 0.5 mg/kg of amphetamine (0.5-Soc.) and one group receiving 2.0 mg/kg of the drug (2.0-Soc.). Administration of saline or drug was intraperitoneal with injection volume kept below .40 cc. for all animals.

The behavior of each animal in the various groups was observed and recorded during four 5-minute periods. Observations were recorded on the Stenograph machine at the rate of one per second. In the initial, predrug observation period, the animal to be observed was placed in the box and observations were made 30 minutes later. If the animal was in a social group, the stimulus rats were placed in the opposite compartment at the same time.

Approximately 24 hours after being observed in the predrug condition, an animal was observed during three 5-minute periods—30, 60, and 90 minutes following administration of amphetamine or saline. Immediately following injection of the amphetamine or saline, the animal was placed in the observation box and, if the animal was in a social group, the three stimulus rats were placed in the opposite compartment of the box. The first observation period started 30 minutes after treatment. The animal was left in the observation box during the intervals between observation periods. By staggering the injection schedule it was possible to take complete records from four animals, one in each observation box.

The observation schedule was such that each of the four observation boxes contained an animal from a different group. Thus, the first observation session, of three observation periods, involved an animal from group S-Sol., 0.5-Sol., 2.0-Sol., and S-Soc.; the second observation session included animals from groups 0.5-Soc., 2.0-Soc., S-Sol., 0.5-Sol., etc. Rats used as stimuli were changed following each observation session. This insured a relatively high level of activity on the part of the stimulus rats during the periods that observations were being made. Three observation sessions, each of which involved complete postinjection observations of four animals, were completed each day.

Drug and saline administration was done by the observer. A "blind" technique was utilized in that the observer did not know whether he was administering saline or a drug concentration. The observer, who was a paid student, was also naive as to the design and objectives of the study.

Five behavior categories that encompassed all of the behavior demonstrated by the observed animal were utilized. The categories were:

a. Locomotion—recorded whenever an animal moved its entire body from one location in the observation box to another location. This included either walking or running, depending upon the speed at which the animal shifted location.

b. Barrier-directed behavior—this category of behavior involved any behavior on the part of the animal that was directed toward the hardware cloth barrier separating the observation box into two compartments.

c. Sniffing—recorded whenever the animal remained in one location in the observation box but was engaged in sniffing as shown by rapid movements of the whiskers, rotational movements of the head and neck, or rearing on the hind legs. The primary criterion distinguishing this category from the inactivity category was some movement of the animal's head. When the rat was engaged at sniffing at the barrier it was recorded as barrier-directed behavior.

d. Grooming—when the rat was engaged in licking, scratching, or biting the fur on some part of its body, this category was recorded.

e. Inactivity—recorded when the animal was lying down in a motionless state with abdomen resting on the floor of the box or when the animal was sitting motionless.

In addition to the above categories, another response measure was obtained from the data. This was the frequency of shifts in behavior shown during an observation period. A behavior shift involved the animal's switching from one of the behavior categories to another and, as pointed out by Bindra and Baran (1), is a measure of the instability in the over-all behavior of an animal.

Animals in the social groups were observed under one other condition. After all animals had been observed in the conditions described above, the barriers in the observation boxes were removed. Animals in the social groups were again administered saline or amphetamine sulfate and placed in the box with one stimulus rat. In this situation the rats interact freely. A five-minute observation was made 30 minutes after the injection. In this situation, however, only two categories of behavior were recorded. These were:

a. Other-animal involved behavior—recorded whenever the observed rat was in physical contact with the stimulus rat.

b. Non-animal involved behavior—recorded whenever the observed animal was engaged in some form of activity that did not involve physical contact with the stimulus rat.

3. *Data Analysis*

Results were described in terms of the frequency of observations recorded in each of the behavior categories during an observation period. In the case of behavior shifts, the total number of shifts demonstrated by an animal during an observation period was utilized. Differences between the performance of *S*s in the various groups were tested for significance by means of the Mann-Whitney U-test, and the null hypothesis was rejected at $p = .05$ or less.

C. RESULTS

1. *Barrier-directed behavior*

Figure 1A shows the mean number of observations recorded in the barrier-directed category for the groups during each of the four observation periods. Animals in the social groups engaged in more of this type activity than animals in the solitary groups. Administration of amphetamine sulfate to animals in the social groups differentially affected this response. Rats in the 2.0-Soc. group engaged in barrier-directed behavior significantly more frequently than those in the 0.5-Soc. group and the S-Soc. group. Although the 0.5-Soc. group showed a higher level of activity in this category than the saline treated group, the differences were significant only at the + 30 minute observation period.

Animals in the solitary groups demonstrated relatively little barrier-directed activity. Although animals in the 2.0-Sol. group engaged in this activity more than those in the other groups, the differences were sot significant.

2. *Sniffing*

Administration of amphetamine tended to increase the sniffing scores of animals in the 0.5-Sol. and 2.0-Sol. groups and to decrease these scores in the 0.5-Soc. and 2.0-Soc. groups. (See Figure 1B). Sniffing scores for animals in the S-Sol. and S-Soc. groups did not differ significantly while scores for the 0.5-Soc. and 2.0-Soc. groups were significantly lower than the scores of the 0.5-Sol. and 2.0-Sol. groups during all postinjection observation periods. These scores for animals in the 2.0-Soc. group were significantly lower than those of the 0.5-Soc. group during the + 30 minute observation period and significantly lower than the S-Soc. group during the + 30 and + 60 minute observation periods. The differences between scores of the 0.5-Soc. and S-Soc. groups were not significant.

The scores of the animals in the 2.0-Sol. group were significantly higher than the scores of the S-Sol. group during all postinjection observation periods. Differences shown between the 2.0-Sol. and 0.5-Sol. groups and the 0.5-Sol. and S-Sol. groups were not significant.

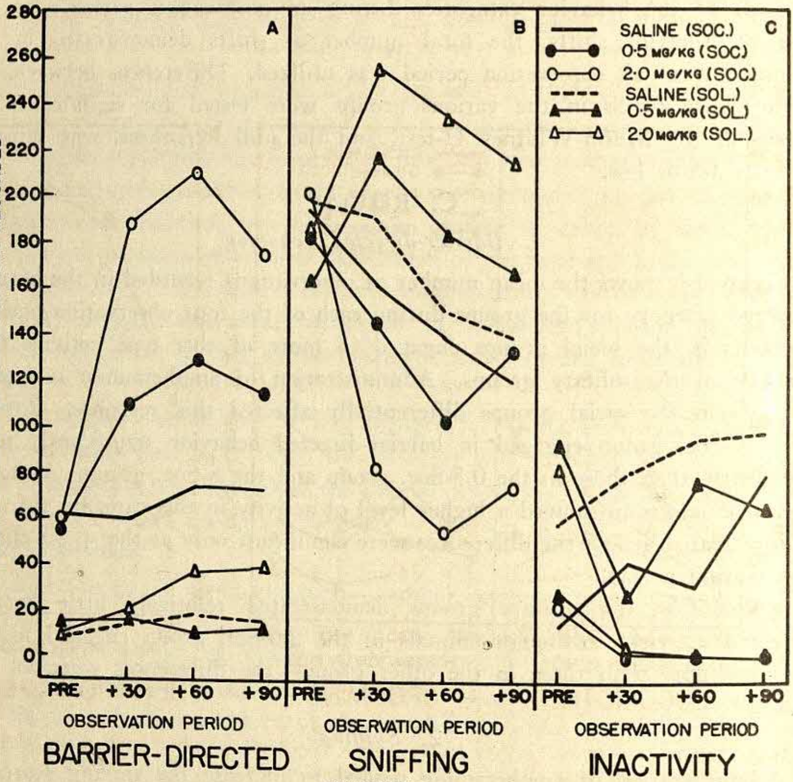


FIGURE 1

MEAN NUMBER OF OBSERVATIONS RECORDED IN BARRIER-DIRECTED, SNIFFING, AND INACTIVITY CATEGORIES OF BEHAVIOR DURING OBSERVATION PERIODS FOR *Ss* IN EACH OF THE SIX GROUPS

3. *Inactivity*

As Figure 1C indicates, inactivity scores increased during the course of the observations in both saline groups. Animals in the S-Sol. group tended to be inactive more than the animals in the S-Soc. group. However, the differences between these groups were not significant. No observations were recorded in the inactivity behavior category for animals in either the 2.0-Soc. or 2.0-Sol. groups or for animals in the 0.5-Soc. group during the postinjection observation periods. Considerable inactivity was recorded for the 0.5-Sol. group. Inactivity scores for this group were significantly higher than those of the 0.5-Soc. group in all three postinjection observation periods.

4. *Locomotion*

Amphetamine treated animals in the social and solitary groups tended to engage in more of this type activity than saline treated animals. (See Figure 2A.) Differences between the locomotion scores for the saline groups were not significant nor were the differences between the scores of the 2.0-Soc. and 2.0-Sol. groups significant. However, during the + 60 minute observation period, the difference between locomotion scores for the 0.5-Soc. and 0.5-Sol. groups was significant with the 0.5-Soc. group showing the highest scores.

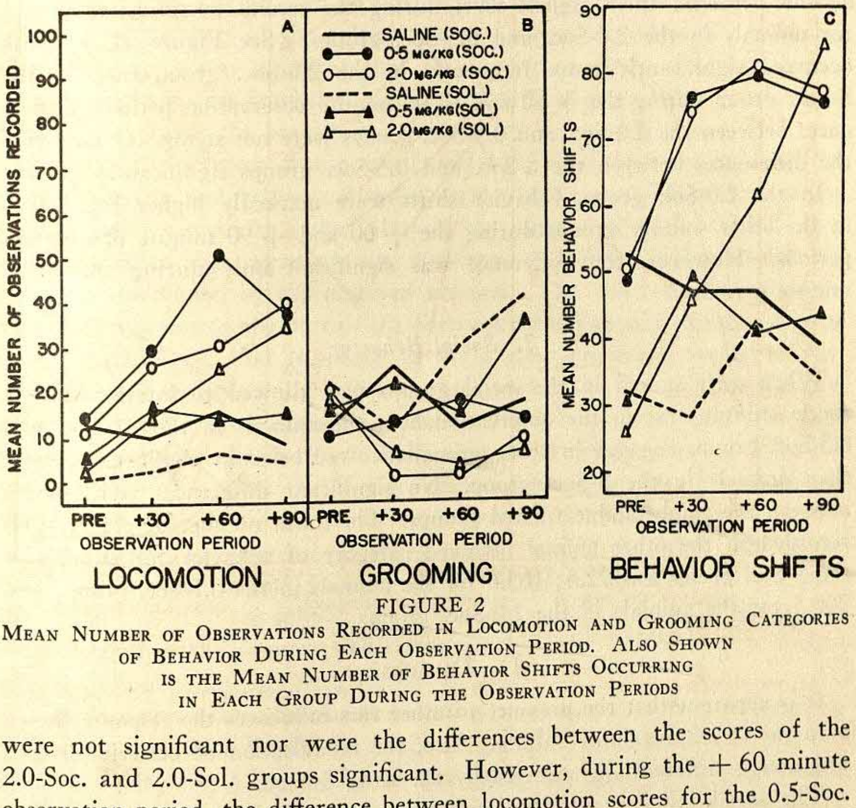


FIGURE 2

MEAN NUMBER OF OBSERVATIONS RECORDED IN LOCOMOTION AND GROOMING CATEGORIES OF BEHAVIOR DURING EACH OBSERVATION PERIOD. ALSO SHOWN IS THE MEAN NUMBER OF BEHAVIOR SHIFTS OCCURRING IN EACH GROUP DURING THE OBSERVATION PERIODS

5. *Grooming*

It is apparent from Figure 2B that grooming scores are reduced by administration of 2.0 mg/kg of amphetamine but not greatly affected by the smaller dosage. During the second postinjection observation period, animals in both the 2.0 mg/kg groups had significantly lower grooming scores than animals in all the other groups and significantly lower scores than the 0.5-Sol. and S-Sol. groups during the + 90 minute postinjection period. As the figure shows, grooming behavior was quite variable.

6. *Behavior Shifts*

Behavior shifts, which were changes from one category of behavior to another, reached their highest level during the second postinjection period for animals in the 2.0-Soc. and 0.5-Soc. groups. (See Figure 2C.) Shifts occurred significantly more frequently in the 2.0-Soc. group than in the S-Soc. group during the + 60 and + 90 minute observation periods. Differences between the 2.0-Soc. and 0.5-Soc. groups were not significant nor were the differences between the S-Soc. and 0.5-Soc. groups significant.

In the 2.0-Sol. group, behavior shifts were markedly higher than those in the other solitary groups during the + 60 and + 90 minute observation periods. However, this difference was significant only during the + 90 minute period.

7. *Social Behavior*

When each animal in the social groups was allowed to interact with a single stimulus rat in the observation box, the animals in the 2.0-Soc. and 0.5-Soc. groups engaged in other animal involved behavior significantly more than animals in the S-Soc. group. No significant differences were shown between the amphetamine treated groups. The mean number of observations recorded in the other animal involved category of behavior for animals in the S-Soc. group was 92.5; 165.6 for the animals in the 0.5-Soc. group; and 200.5 for the animals in the 2.0-Soc. group.

D. DISCUSSION

It is apparent that the presence of other rats influences the response shown by a rat to amphetamine sulfate. Also, the modification of behavior in this situation is differentially affected by the drug dosage.

The social influence of drug response was most obvious in the behavior described as barrier-directed. Originally, the barriers had been included in the observation boxes in order to facilitate the recordings. In previous work (5, 6, 7), observations had been made on pairs of animals but observing and recording the behavior of an animal in a larger group is somewhat cumbersome. By placing a hardware cloth barrier in the box, it was felt that observations of the treated animal could more readily be made and that animals on the opposite side of the barrier would provide adequate social stimulation. A preliminary investigation, however, indicated that behavior directed at the barrier separating the observed animal from the stimulus rats was engaged in by amphetamine treated rats during a large part of an observation period. Consequently, this type behavior was included as a behavioral category in the study.

Although the category of barrier-directed behavior was defined as any behavior on the part of the animal that was directed toward the hardware cloth barrier, it was apparent that the barrier, as such, did not elicit a great deal of response from the animals. It might be argued that much of the behavior involving the barrier was exploratory or manipulatory. However, animals in the solitary groups spent relatively little time engaged in this type of activity and behavior directed at the barrier was no more common than behavior directed at other parts of the compartment. In the situation where stimulus rats were present, however, a variety of barrier-directed behavior was shown by the observed animals. The most common form involved an attempt on the part of the observed animal to sniff, groom, or even bite a stimulus rat that happened to be in close proximity to the barrier. The observed rats generally followed the stimulus rats by changing position along the barrier, even when the stimulus rats were on the side of the compartment opposite the barrier. Other forms of barrier-directed behavior included rearing against the barrier, climbing on it, and biting or sniffing at the mesh. It was obvious, however, that nearly all barrier-directed behavior was due to the presence of the stimulus rats on the other side of the barrier. That the barrier-directed activity scores might agree closely with the amount of time an animal would actually spend in social behavior was demonstrated in the situation where the barrier was removed and a treated animal allowed to interact with a stimulus rat.

The increase in barrier-directed behavior as a result of amphetamine sulfate administration would not have been predicted on the basis of previous research reported by Chance (2). In this study, the behavior of mice treated with amphetamine and placed in an aggregate was described. Much of the behavior demonstrated by the mice in this situation appeared to be a type of avoidance behavior. Chance states that "Such mice, when placed together in groups, react to one another's squeaks by little jumps and by adopting a position of readiness for action. . . . If, moreover, they are approached by other mice, they either attempt to run away or adopt a defensive attitude towards the approaching animals." It is doubtful that any of the behavior shown by amphetamine treated rats in their barrier-directed activity could be classed as avoidance. It suggests, rather, the opposite. An earlier study by Heimstra (5) showed that pairs of rats treated with 0.5 mg/kg of amphetamine sulfate engaged in significantly more nonaggressive social behavior than saline treated rats or pairs of rats in which one animal had received the drug. This also suggests that amphetamine administration, at least light or moderate dosages, does not result in avoidance behavior in a social situation.

In the studies reported by Chance (2, 3) and by Gunn and Gurd (4), a marked increase in activity took place when amphetamine treated mice were placed in a group situation. Animals tended to rush about rapidly and display a considerably higher level of activity than mice with similar dosages but housed individually. In the present study, the locomotion scores, which included running and walking, were higher for animals in the treated social groups than those in the solitary groups. However, the only significant difference in locomotion scores shown between social and solitary groups involving amphetamine treated animals was between the 0.5 mg/kg groups during the + 60 minute postinjection observation period. The difference in activity level in these groups was also shown in the inactivity scores where the solitary group had significantly higher scores during all postinjection observation periods.

Another category of behavior that showed marked differences between amphetamine treated social and solitary groups was the category of sniffing. Sniffing, which can be considered as a component of general activity, has been found to be a sensitive indicator of drug effects (1). In the present investigation, sniffing scores for the amphetamine treated solitary groups were significantly higher than scores of the social groups. In comparing the figures showing the frequency of occurrence of barrier-directed behavior and sniffing for the social groups, it can be seen that sniffing decreased as barrier-directed behavior increased.

The results of this investigation support previous studies which have shown that group conditions modify response to amphetamine sulfate. Although the earlier studies (2, 3, 4) differed from the present investigation in that all animals in a group situation were drugged and were allowed to interact freely, the present study showed that the presence of nontreated rats, placed behind a barrier, would also influence a rat's response to amphetamine. Whether social influence is particularly apparent in the response to CNS stimulants or whether it modifies response to other types of drugs will be determined in following studies in this series.

E. SUMMARY

Six groups of eight rats each were utilized. Animals in three of the groups were observed alone in an observation box (solitary condition) while three groups were observed in a social situation with stimulus rats present in a second compartment of the box. One social and one solitary group received saline, one of each of the groups received 0.5 mg/kg of amphetamine sulfate, and the remaining two groups received 2.0 mg/kg of amphetamine. Each rat was observed during four 5-minute observation periods—a pretreatment period, a + 30, + 60, and + 90 minute postinjection period. Five categories of behavior were observed and recorded. These included (1) behavior directed at the barrier separating the observed rat from the stimulus rats, (2) locomotion, (3) inactivity, (4) grooming, and (5) sniffing. The number of times an animal shifted from one category of behavior to another was also analyzed.

1. Amphetamine treated animals in the social groups tended to engage in more barrier-directed behavior than saline treated animals.

2. The activity level of amphetamine treated social groups was higher than the level demonstrated by animals in the amphetamine treated solitary groups although these differences were significant only for the 0.5 mg/kg groups at the + 60 minute observation period.

3. Administration of amphetamine sulfate increased the sniffing scores for animals in the solitary groups and decreased the scores of animals in the social groups.

The investigation showed that the presence of other rats modified the response of rats to amphetamine sulfate.

REFERENCES

1. Bindra, D., & Baran, D. Effects of methylphenidylacetate and chlorpromazine on certain components of general activity. *J. Exp. & Anal. Behav.*, 1959, **2**, 343-350.
2. Chance, M. R. A. Aggregation as a factor influencing the toxicity of sympathomimetic amines in mice. *J. Pharmacol. & Exp. Therap.*, 1946, **87**, 214-219.
3. ———. A peculiar form of behavior induced in mice by amphetamine. *Behaviour*, 1949, **1**, 64-70.
4. Gunn, J. A., & Gurd, J. The action of some amines related to adrenaline. Cyclohexylalkylamines. *J. Physiol.*, 1940, **97**, 463-470.
5. Heimstra, N. W. Effects of chlorpromazine and amphetamine on the behavior of rats in a competitive situation. Unpublished Ph.D. dissertation, Univ. of Rochester, 1960.
6. ———. Effects of chlorpromazine on food dominance and fighting behavior in rats. *Behaviour*, in press.
7. Heimstra, N. W., & Newton, G. Effects of prior food competition on the rat's killing response to the white mouse. *Behaviour*, 1961, **17**, 95-102.

8. NORTON, S. Behavioral patterns as a technique for studying psychotropic drugs. *In* S. Garattini and V. Ghetti (*Eds.*), *Psychotropic Drugs*. New York: Elsevier Press, 1957, pp. 73-82.
9. NORTON, S., & DEBEER, E. J. Effects of drugs on the behavioral patterns of cats. *Ann. N.Y. Acad. Sci.,* 1956, **65**, (4), 249-257.
10. NORTON, S., & TAMBURRO, J. Effects of hallucinogins on spontaneous behavior patterns of animals. *J. Pharmacol. & Exp. Therap.,* 1958, **122**, 57a.
11. NOWLIS, V., & NOWLIS, H. The description and analysis of mood. *Ann. N.Y. Acad. Sci.,* 1956, **65**, 345-355.

Department of Psychology
University of South Dakota
Vermillion, South Dakota

$20.00 per annum
$15.00 per volume
$7.50 per half volume

QUARTERLY
Two volumes per year
Immediate publication

April, 1962
Volume 53, Second Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

APRIL, 1962

(Manuscripts are printed in the order of final acceptance)

The perception of faces—a further study of the Engel effect 245
By William H. Ittelson and Bernard Seidenberg

Chemical studies of behavior: VI. Placebo and dramamine as methodological controls, and effects on moods, emotions and motivations 257
By G. R. Wendt, Jean S. Cameron, and Priscilla G. Specht

An exploration of some personality correlates to business risk-taking behavior . 281
By Joseph G. Phelan

A study of test-retest results obtained on aircrew proficiency examinations . 289
By Kenneth S. Teel and Robert L. Petersen

The stressed conditioning of psychotics 295
By James K. Feibleman

Conditioned responses: fact and artifact 301
By Sam J. Korn and Livingston Welch

(OVER)

Copyright, 1962, by The Journal Press
Provincetown, Massachusetts, U. S. A.
Entered as second-class matter January 26, 1937, at the post-office at
Provincetown, Mass., under the Act of March 3, 1879
Second-class postage paid at Provincetown, Mass.

Symboling: a kind of behavior 311
By Leslie A. White

Brightness of monochromatic light in scotopic and photopic vision . . . 319
By Gösta Ekman and Teodor Künnapas

What psychologists should know about marketing and advertising research . 329
By Frederick J. Thumin

Religious attitudes at a small denominational college as compared with Harvard and Radcliffe 349
By Harry K. Wells

Hypno introspection: a contribution to the theory of hypnotherapy. Part I . . 383
By Seymour Halpern

Personal beliefs of schizophrenics and their parents 387
By Leonard W. Krinsky

Perceptual threshold decrements as a function of skill and chance expectancies . 399
By E. Jerry Phares

Some personality correlates of verbal conditioning 409
By W. Robert Hetrick and Kurt Haas

Stimulus generalization with aggressive verbal stimuli: a new paradigm . . 417
By Arnold H. Buss

Anticipatory behavior: temporal matching by normal and psychotic subjects . 425
By H. E. King

Abstract conditioning 441
By Jerry E. Waters and Frank Kodman, Jr.

Stereotypes of the scientist as seen with the Gordon Personal Profile and Gordon Personal Inventory 453
By John R. Braun

The influence of "fake" personality evaluations of self description . . . 457
By Gloria Carey Evans

Food intake as a function of duration of food deprivation 465
By Robert H. Dufort and John H. Wright

Verbal concept attainment: a function of the number and strength of positive instances presented 469
By M. S. Mayzner and M. E. Tresselt

The effect of convergence on perceived size and distance 475
By Walter C. Gogel

The effect of systematic and nonsystematic presentation of stimuli on peripheral visual acuity 491
By Harry W. Karn, Lee W. Gregg, and Gordon F. Pitz

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 247-255.

THE PERCEPTION OF FACES—A FURTHER STUDY OF THE ENGEL EFFECT*[1]

Department of Psychology, Brooklyn College

William H. Ittelson and Bernard Seidenberg

A. Purpose

Engel (1, p. 303) has reported an interesting stereoscopic effect. "A pair of photos depicting two different faces serve as the two monocular targets in a stereoscopic device. The faces are similar in size and position of facial parts and, by means of a previous adjustment of the apparatus, stimulate corresponding places on the two eyes of an observer. At the outset of the presentation only one of the monocular targets is illuminated. The observer, viewing binocularly, is asked to describe what he sees. The illumination is then extinguished and a second presentation given, but this time the second monocular target is also illuminated but to a very slight degree. The observer is then asked to state whether any change has taken place in what he sees. Following this report, the illumination is again extinguished and the procedure repeated, but again adding a slight increment of illumination to the second target while keeping the first at the same level. The entire procedure is repeated until the second target reaches the illumination level of the first. The remarkable thing is that, although the second target is raised in illumination, at each step the observer reports that he sees the same face as before. This continues to be his report even after the second target reaches the same degree of illumination as the first."

"The reverse procedure is then begun. The second target is kept at the same level of illumination while the illumination of the first is brought down in small decrements on each trial. On the last trial only the second target is illuminated. What is most striking is that the observer will continue to report seeing the same face to the bitter end. This occurs despite the fact that on the last trial a clearly different face is shown than on the first."

The purpose of this study was to establish the generality of the Engel effect and to explore some of its aspects. Certain procedural variations were

---

* Received in the Editorial Office on October 16, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported by research grant M2503 from the National Institute of Mental Health Service in cooperation with the VA Hospital, East Orange, N.J. The authors wish to thank Edward Engel for providing pre-publication results of his investigation and for the many valuable suggestions which he made.

also investigated, such as the effect of the amount of light increment, the effect of repeated trials, and the influence of eye dominance. Still another area of interest involved the similarity of the pairs of pictures to be used. To study the effect of similarity, two sets of disparate photos, a white male–white female (WM-WF) set, and a white male–Negro male (WM-NM) set were included as stimuli along with the three sets of white male–white male (WM-WM) photos.

B. METHOD

1. *Design*

A factorial design was employed consisting of four groups of 12 *S*s each. The three experimental groups were exposed first to three sets of similar white male faces followed by a disparate pair of photos, either the WM-WF or the WN-NM set. These three groups differed from each other only in the size of the illumination steps used and were designated the small interval (SI), medium interval (MI), and the large interval (LI) groups. The point of equal illumination, which also served as the initial level of illumination for picture A and the final level for picture B, was 15 foot candles. The ascending illumination steps, beginning with essentially zero illumination, represented foot candle increment ratios of 1.45(SI), 1.70(MI), and 2.50(LI) of each setting over the previous one. The descending steps, beginning with 15 foot candles, represented ratios of .69(SI), .59(MI), and .40(LI) of the immediately preceding value.

A control group was run at the smallest increment level (SI) and differed from the three experimental groups in that a disparate pair of photos, either the WM-WF or the SM-NM set was exposed to the subjects first, followed by the three similar pairs.

One picture of each pair was designated picture 1 and the other picture 2 (in the disparate pairs, picture 1 was the white male in each case). Picture 1 was always illuminated at the beginning of a series. The three sets of similar photos, sets A, B, and C were presented in all six possible orders twice within each of the experimental and control groups. In the first block, picture 1 of each set was exposed to the right eye of the subject, while in the second block, picture 1 was exposed to the left eye of the subject. This procedure controlled for the effects of eye dominance and order within each group.

2. *Subjects*

The *S*s consisted of 48 hospitalized veterans from the medical service of a VA hospital. Though suffering from a wide assortment of medical ailments, all were in a convalescent stage and were due for discharge within a week or two. They ranged in age from 18-48 years and from eighth grade to some college in education.

Patients were assigned consecutively to random ordered conditions. An exception involved the 13 Negro patients who were assigned to the first available condition involving the WM-WF disparate pair of photos only.

3. *Procedure*

Each *S* was seen individually and was asked to volunteer for a study of "visual perception, or how people see things; particularly faces of people." *S*s were then given an eye test using the Bausch and Lomb Orthorater which yields measures of near and far acuity for both eyes and for each eye separately, vertical and lateral phoria, and stereopsis. All *S*s whose corrected vision for both eyes was worse than 20/60 or where the discrepancy between the two eyes was more than 10 (minimum case, 20/60—20/50) were eliminated from the study.

*S*s who met these initial criteria were then measured for interocular distance (ID) and seated before the stereoscope (an American Optical Co. Troposcope). Preliminary adjustments for ID and height were made, as were finer adjustments by means of a set of test slides (star-circle pair). The settings were then confirmed with another pair of slides (airplane-hangar pair).

At this point the *S* was given the following instructions: "We are now going to start. I'm going to show you pictures of people's faces and frequently ask you if you see a change. If you do not, say 'No' and we'll go on. If you do, describe the change that you saw. Try to keep your head still and look straight ahead with both eyes open. Try not to stare. If you get tired, tell me and we can take a break after we finish this run. Now close your eyes." At this point the first pair of photos was inserted in the troposcope. "Please open your eyes! What do you see?" Here, the *S*'s description of the face was recorded and the first increment was then made. "Any change?" The *S*'s response or lack of it was recorded verbatim at each step until the entire procedure was completed. If the *S* reported a new face, this was recorded but the procedure was completed in any event. At the end, whether the *S* reported a new face or not, he was asked, "Is this the person we started with?" and his response was recorded. The procedure was repeated for the other trials, a total of four times for each *S*.

The present procedure differs from Engel's in that the subjects viewed the stimulus field continuously. Since pilot work indicated that it was not necessary to extinguish the illumination of the stimulus field between reports, this simplification of the procedure was adopted.

The scoring measures used are based on response categories since, in the majority of cases, a scoreable response did not occur till the end of the procedure. The categories were Same (S), used in the case where the subject perceived no change in the photos throughout the procedure; Different (D), used when the subject indicated that a change had been experienced only at the very end of the procedure; and Different Before End (DBE), used when a subject indicated that he had experienced a change sometime before the end of the procedure.

C. RESULTS

In Table 1, the frequency of S, D, and DBE responses (three per subject) to the similar pairs are given for the three experimental groups and the control group. These results demonstrate conclusively that the effect, as Engel reports it, is present and widespread in the population tested. Sixty-two per cent of the 108 similar pair trials in the experimental group ended in the subject reporting no change, while in only 20 per cent of the trials was a change reported before the end of the procedure. Although these proportions are essentially reversed in the control group, even here 39 per cent of the trials ended in a report of no change.

TABLE 1
FREQUENCY OF RESPONSES FOR ALL GROUPS FALLING INTO EACH RESPONSE CATEGORY FOR SIMILAR PHOTOGRAPHS

Response categories	Experimental groups SI	MI	LI	Total	Control group SI	Grand total
S	22	22	23	67	14	81
D	5	5	9	19	4	23
DBE	9	9	4	22	18	40
Total	36	36	36	108	36	144

There is, however, strong evidence of individual differences in response. Table 2 shows the number of subjects giving different combinations of response to the similar photos.

In the experimental groups, 64 per cent of the subjects (23) reported either no change at all or changes only at the end of the procedure in all three presentations while 36 per cent (13) reported one or more DBE response out of the three trials. In the control group such individual differences are also manifest in that, while 75 per cent of the control subjects gave at least one DBE response, 25 per cent gave no DBE response at all.

TABLE 2

FREQUENCY OF SUBJECTS GIVING VARIOUS PATTERNS OF RESPONSE TO THREE SIMILAR PAIRS FOR ALL GROUPS

Response patterns	Experimental groups SI	MI	LI	Experimental total	Control	Grand total
3S	5	4	5	14	2	16
2S, 1D	1	1	2	4	1	5
2S, 1DBE	0	2	1	3	2	5
1S, 2D	1	2	1	4	0	4
1S, 1D, 1DBE	2	0	1	3	1	4
3D	0	0	1	1	0	1
2D, 1DBE	0	0	0	0	1	1
1S, 2DBE	2	2	0	4	0	4
1D, 2DBE	0	0	1	1	0	1
3DBE	1	1	0	2	5	7
Total	12	12	12	36	12	48

Tables 1, 3, and 4 show the frequencies of response of the various groups in terms of size of illumination change interval, eye dominance, order of presentation of the pairs of photos, and the three pairs of photos respectively. The proportion of responses in each scoring category for the three experimental groups, varying only in interval level, did not differ significantly as determined by a chi square test. Another chi square test failed to yield a significant difference in response between subjects who had the first picture of a pair exposed to the left eye and those who had the first picture exposed to the right eye.

No significant difference was obtained in terms of differential response to the three pairs of similar photos. However, there is a slight but not significant tendency for the proportion of DBE responses to increase from the first to the third trial. If only the responses to the first and third trials are compared, the chi square of 4.76 with 2df that is obtained is significant at the 9 per cent level. While this result is not significant, it suggests that some change in performance toward a greater sensitivity to change is occurring.

Table 1 indicates that a major difference in performance occurred between the experimental groups and the control group. These groups differed in that the experimental groups were first exposed to the three similar pairs of photos followed by the one disparate pair, while in the control group, the disparate pair preceded the three similar pairs. The proportion of response differs greatly between the two groups, this difference yielding a $\chi^2 = 11.63$ which for 2df is significant beyond the .01 level. The greater tendency for the control group to give significantly more DBE responses is probably

TABLE 3

FREQUENCY OF RESPONSE IN EACH CATEGORY AS A FUNCTION OF EYE FIRST EXPOSED

Response Categories	Experimental groups SI		MI		LI		Experimental total		Control group SI		Grand total	
	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
S	12	10	10	12	13	10	35	32	10	4	45	36
D	2	3	3	2	3	6	8	11	2	2	10	13
DBE	4	5	5	4	2	2	11	11	6	12	17	23

directly attributable to the influence of prior exposure of the disparate photos in establishing a set of expectation on the part of the control subjects to perceive changes in the similar pairs.

TABLE 4
FREQUENCY OF RESPONSE OF THE EXPERIMENTAL GROUPS COMBINED AS A FUNCTION OF SIMILAR PICTURE PAIRS AND OF PICTURE TRIAL

	Trial			Pair		
	1st	2nd	3rd	A	B	C
S	24	24	19	20	22	25
D	8	4	7	8	4	7
DBE	4	8	10	8	10	4

A rather obvious but interesting difference was found with respect to the performance of both the experimental and control groups *vis-à-vis* the similar and disparate photos. The data in Table 5 indicate that in every case, in both the experimental or control groups, a change was reported in response to the disparate pairs. Indeed, only two D responses were given, the remainder being DBE. This difference is significant beyond the .001 level and supports the rather obvious notion that the Engel effect breaks down as the stimulus pairs become more disparate.

TABLE 5
FREQUENCY OF RESPONSES OF EXPERIMENTAL AND CONTROL GROUPS SIMILAR AND DISPARATE PHOTOS

	Experimental group		Control group	
	Similar	Disparate	Similar	Disparate
S	67	0	14	0
D	19	2	4	0
DBE	22	34	18	12

D. DISCUSSION

The results of this investigation strongly support the existence of the Engel Effect as a perceptual phenomenon. The importance of this is heightened when it is realized that the pairs of photographs labeled "similar" are only similar in a comparative sense. Actually, each pair consisted of pictures of two distinctly different and easily distinguishable young, white males. Then, too, the procedure permitted constant observation by the subject. While it might be assumed that the issue of constant observation might have no bearing on the results, or even that such observation might enhance the effect, it must also be noted that particularly in the LI group, rather large increments and decrements in the illumination level of the two pictures were being made right before the subject's eyes.

That the Engel effect is not an artifact of the procedure is supported by two additional findings. First of all, while the effect was demonstrated in the great majority of the trials in the experimental group, it was ascertained that some individuals seem to be more sensitive to the perception of change in the situation than are others. Secondly, the performance of the control group differed markedly from that of the experimental group, 50 per cent of the responses falling in the DBE category. It is interesting to note, however, that the attenuation of the effect in the control group occurred only after the control subjects were exposed to the rather dramatic change involved in the disparate photo situation. If a change of set or expectancy is involved, it is surprising that no greater significant tendency toward the perception of change was found in the second and third similar photo trials in the experimental group. On the basis of the findings presented, the Engel Effect would seem to fit into the general pattern of results from many other studies which illustrate that the observer has a wide latitude in the interpretation of percepts.

There is, however, an important stimulus limitation. When sufficiently disparate stimuli are used, the Engel Effect breaks down. While there are, in this case, no individual differences in terms of whether or not change is finally reported, since changes were reported by all the *S*s involved in the present study, such differences are apparent in terms of when in the procedure change is reported.

E. Summary

Two similar but distinct faces are presented, one to each eye, by means of a stereoscope. Initially, one face is fully illuminated while the other is completely dark. The dark face is gradually illuminated until it equals the illumination of the first face which is then gradually decreased till it is darkened. Subjects frequently report no change in the face being observed. The Engel Effect occurs when the *S* reports no change being observed.

Thirty-six male *S*s were divided equally into three experimental groups differing only in the size of the illumination change interval used. Each subject was exposed to four trials, the first three involving "similar" pairs of photos and the fourth trial involving a disparate pair, either a male-female or a white male-Negro male pair. The effects of eye dominance and order of presentation of the similar pairs were controlled. A control group consisting of twelve *S*s differed from the first three groups only in that the disparate photos were presented prior to the three sets of similar photos.

In 62 per cent of the similar pair trials in the experimental groups no change was reported, thereby confirming previous pilot work. In the control group, however, no change was reported in only 39 per cent of the trials.

The results indicate the following:

1. The report of no change following the end of the procedure (Engel Effect) occurred in 62 per cent of the trials in the experimental group.

2. While no change is seen in 39 per cent of the similar pair trials in the control group, there is a significant difference between the experimental groups and the control group in the frequency of reports of no change.

3. There is a highly significant difference in the response to similar as opposed to disparate pairs of photos, indicating that the Engel Effect is dependent, in part, on the relative similarity of the photos involved.

REFERENCE

1. ENGEL, E. Binocular methods in psychological research. *In* Kilpatrick, F. (*Ed.*), *Explorations in Transactional Psychology*. New York Univ. Press, 1960.

Department of Psychology
Brooklyn College
Brooklyn 10, New York

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 257-279.

CHEMICAL STUDIES OF BEHAVIOR: VI. PLACEBO AND DRAMAMINE AS METHODOLOGICAL CONTROLS, AND EFFECTS ON MOODS, EMOTIONS AND MOTIVATIONS*

Department of Psychology, University of Rochester

G. R. WENDT, JEAN S. CAMERON, AND PRISCILLA G. SPECHT[1]

A. INTRODUCTION

This paper is the second of a series describing experimental results of various drug studies. The first paper was introductory in that it described our methodology, giving objectives, rationale and our choice of solutions in undertaking the investigation of drug effects on human subjects (13). We wished to obtain a descriptive picture of the effects of particular drugs on moods, emotions and motivations uncomplicated by the effects of suggestion, expectation, conditioning, excitement, etc. The methods were sensitive and yielded good reliability of results.

This present paper has two purposes: first, to describe our use of both placebo and dramamine (dimenhydrinate) as methodological controls in drug research; second, to describe their effects on emotional and motivational states, when procedures were carefully designed to minimize placebo effects and to promote validity and sensitivity of measures of drug effects.[2]

By methodological control we mean the inclusion of placebo and dramamine in an experimental design with other drugs using these two as standards for evaluation of results describing the effects of the other drugs. The rationale for use of two such standards is that data from placebo yield an assessment of the degree of intrusion of unwanted or uncontrolled factors; data from a thoroughly familiar drug (as dramamine is for these investigators) evaluate the sensitivity and validity of procedures. If the data assess dramamine accurately, we then know that the methods are yielding the same results as in the past. Thus the probability is greatly increased that the data on new drugs are meaningful to the extent that they have reportable effects.

---

* Received in the Editorial Office on November 6, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The four researches here reported were supported by a number of agencies in addition to the University of Rochester. The main support came from G. D. Searle and Co., makers of DRAMAMINE. But since we were simultaneously investigating effects of other drugs, support over the period covered came also from the Office of Naval Research (Contract No. N6ori-126, T. O. I.), the National Institute of Mental Health (Grant No. M-681), Wallace Laboratories and the Woodward Research Corporation. The personnel was as follows: 1954: Jean S. Cameron, G. R. Wendt; 1955: Jean S. Cameron, Sallie Laverty, G. R. Wendt; 1958: Jean S. Cameron, Eugene Sachs, G. R. Wendt; 1959: Jean S. Cameron, Priscilla G. Specht, G. R. Wendt.

[2] The present analysis and presentation of data omits much of the material we have at hand because the purposes of this paper are limited. Data on individual differences in response, correlates of adjective score changes, and others (13, p. 174) are deferred for subsequent papers.

Our familiarity with dramamine is based on the fact that it has been informally or formally studied in our laboratory for its efficiency as an antihistamine, motion sickness preventive, anti-emetic, hypnotic, sedative and tranquilizer. This paper describes the last four of seventeen studies of its effects as a sedative and tranquilizer.

A distinction needs to be made at this point between our data obtained after placebo and the kinds of changes described as "placebo effects" in much of the literature. Such placebo effects are apparently regarded by most as ubiquitous (2, 8). The nature of placebo effects has been sketched by us (13, pp. 198-200). These include a variety of effects arising from *S*'s past conditioning, expectation, undesirable attitudes, needs, etc. We were willing to accept tentatively that they are inevitable in drug work with patients (12); their emotional needs are usually too strong to allow the experimenters to generate unbiased attitudes in the *S*s. By contrast, in the four experiments on college student *S*s here described, such placebo effects were minimal or absent. We are thus able to speak of placebo "data" rather than placebo "effects." This is true of 9 out of 10 of the placebo determinations here reported; the data after placebo either showed no change or a change attributable to external factors, e.g. lunch, which occurred between taking the lactose capsules and gathering data on their "effects." In the tenth case there were weak placebo effects.

The data presented are from the long form of our Adjective Check List (13, p. 205). The ACLs gave the *S* an opportunity to respond to self-descriptive adjectives (i.e., active, aggressive, ambitious, amused, angry, annoyed, etc.) which sampled mood, emotional and motivational states affected by drugs. Scores were attached to each adjective depending on its rejection, acceptance or strong acceptance by the *S*. The score change from before drug to during drug was the measure of drug effect. The method has considerable *a priori* and empirical validity. Our experience leads us to present the data of each adjective as a more or less valid or true index of the state it describes, rather than to deal with the data without interpretation as "verbal behavior."

The analysis of data in this paper is by individual adjectives (rather than by clusters of adjectives) and consists of *change scores* from before drug to during drug for each adjective. Our evaluation of the data is based on consistency of results obtained in repeated experiments, on internal consistency of the data, on correspondence of results with expectation (e.g., that a small dose will have less effect than a large dose), and on the correspondence of the ACL data to observed behavior (not now reported). Readers who prefer statistical evaluation are urged to make any legitimate calculations from the data presented.

B. PROCEDURES

The data presented in this paper are taken from four experiments conducted in 1954, 1955, 1958 and 1959 with comparable procedures (13). In each experiment male college-student *S*s served for a number of sessions (4 to 15), the sessions being separated by two or more days. On each day the subject came twice, once in the late morning, when he filled out report forms, completed an ACL and took his medication; again two hours after medication, to report on the two "outpatient" hours and to complete ACLs. The method was modified double blind (13, p. 197). A number of drugs and placebo were included in each study, different *S*s being on different drug sequences so as to counterbalance variable factors. *S*s came to the laboratory individually. *Every S in every experiment completed all scheduled sessions*, a good indication of their involvement and interest.

Matters specific to each study (i.e., unique or different from the procedures previously described, 13) were as follows:

> 1954, May 17-May 27. Twenty-four *S*s served for four sessions during which each received placebo, 50 mg of dramamine, 5 mg of profetamine phosphate, and the combination of dramamine and profetamine. Each of the 24 *S*s was placed on a different one of the 24 possible sequences of dose. A 200 word ACL was used, but only the data of those 124 adjectives in common with the present 130 word list are presented in detail. The *S*s were allowed *four* alternative responses to each word; crossing out the adjective if it definitely did not apply; doing nothing if it did not apply; a single check if it did apply; a double check if it strongly applied. These alternatives were scored 1, 2, 3, and 4 respectively. The *S*s did not have two initial training days as previously described (13, pp. 194-5) because they had just completed 175 hours or more of similar experimentation using the ACL.
>
> 1955, May 2-June 1. Twenty-two *S*s served for eight sessions each. The doses used were placebo (twice), 100 mg of dramamine, 5 mg of dexedrine, the combination of dramamine and dexedrine, and three dose levels (10, 15 and 20 mg) of the experimental drug, SC-2491.[3] (The

---

[3] G. D. Searle and Co., P. O. Box 5110, Chicago 80, Ill.

last-named proved to be inactive on the ACL except at 20 mg where the only significant effect reported was headache.) The counterbalancing of drug sequence has been described (13, pp. 194-6). The 130 word ACL was used with response alternatives and scoring as in the 1954 work. No initial training days were required because all *S*s had at least 175 hours of prior training. Each *S* came *three* times on each day: a session before drug and sessions at two hours and at four hours after drug. The *S*s came to a large second floor office rather than the drug research laboratory used in the other studies (13, pp. 180-1, 203).

1958, February 28-April 29. Thirty *S*s served for thirteen sessions each. The doses used were placebo (5 times, two of which were training days), 600 mg of aspirin, 80 mg of dramamine, 600 mg of meprobamate (3 times), 800 mg of meprobamate, 10 mg of amphetamine sulphate, and 600 mg of meprobamate combined with 10 mg of amphetamine. Procedures were otherwise exactly as described in 13. This included the elimination of the cross-out as a response alternative so that the *S*s left blank all adjectives which did not apply. Blanks, checks, and double checks were scored 1, 2 and 3 respectively.

1959, March 23-May 28. Thirty-six *S*s served for fifteen sessions each as shown in 13, Table 1. This included four placebo sessions (two for training) and two sessions with 80 mg of dramamine. (See above reference for other drugs used and for the exact procedures.)

C. RESULTS

1. *Introduction*

An important consideration in evaluating the propriety of using *change scores* for assessment of drug effect is whether the *S*s check approximately the same number of adjectives after drug (or placebo) as before. We have made frequency counts of checking (treating single checks and double checks alike) for the 1958 and 1959 data. These data are remarkably constant for the total group of *S*s, both within and between experiments. The totals for the two treatments are *identical*: the *S*s responded to an average of 30 adjectives before treatment and 29 after treatment, on both placebo and dramamine.

The results of these four experiments, as they relate to our present purposes, are displayed in Tables 1 and 2. Although the methods of 1954 and 1955 were not exactly the same as those of 1958 and 1959, the similarity of results permits the reasonable assumption that they were functionally similar.

We display the dramamine data in Table 1. These data have not been adjusted by subtracting the placebo data displayed in Table 2. The changes after placebo were mostly so small that for practical purposes they can be

TABLE 1

EFFECTS OF DRAMAMINE ON MOODS, EMOTIONS AND MOTIVATIONS

Total *Change Scores* for Each Group of *S*s

(Scores adjusted to N = 30)

Adjective ( ) Ref. No., Table 2	100 mg 1955	80 mg 1958	80 mg '59 I	80 mg '59 II	50 mg 1954	100 mg 4 hr., '55
1 drowsy (30)	+50	+29	+27	+34	+26	+34
2 sluggish (2)	+50	+22	+27	+22	+16	+34
3 weary (34)	+46	+15	+17	+16	+9	+31
4 tired (6)	+45	+19	+22	+15	+12	+30
5 sleepy (3)	+39	+29	+26	+28	+19	+29
6 lazy (7)	+38	+15	+16	+17	+19	+22
7 washed-out (83)	+38	+13	+17	+14	+15	+31
8 dull (49)	+35	+14	+17	+19	+4	+29
9 languid (13)	+35	+4	+21	+17	+14	+20
10 weak (28)	+34	+12	+12	+21	+9	+26
11 quiet (90)	+24	−1	+13	+16	+16	+14
12 lackadaisical (1)	+23	+2	+12	+16	+16	+11
13 drifting (8)	+22	+13	+12	+14	+16	+14
14 tame (91)	+22	+2	+8	+11	+6	+3
15 withdrawn (18)	+22	+9	+10	+12	+2	+15
16 dizzy (47)	+20	+11	+9	+8	+1	+5
17 leisurely (10)	+20	0	+3	+3	+11	+12
18 light-headed (16)	+20	+8	+8	+10	+2	0
19 silent (25)	+20	+12	+14	+20	+8	+16
20 confused (23)	+19	+8	+2	+9	−1	+4
21 detached (65)	+19	+6	+7	+12	+2	+11
22 doubtful (57)	+18	+4	+3	+6	−5	+7
23 indifferent (5)	+18	+9	+11	+14	+4	+11
24 subdued (4)	+18	+9	+13	+19	+6	+11
25 disturbed (79)	+16	0	0	+2		+11
26 dreamy (9)	+16	+4	+8	+14	+12	+11
27 disinterested (29)	+14	+12	+12	+12	+4	+15
28 close-mouthed (50)	+12	+8	+8	+19	+10	+12
29 belligerent (103)	+11	0	−2	+1	−1	+5
30 depressed (26)	+11	+9	+4	+7	+6	+8
31 dubious (66)	+11	+5	−2	+5	+4	+7
32 headache (19)	+10	+2	+4	+5	+2	+7
33 impulsive (127)	+10	−3	+2	−2	−2	+3
34 intoxicated (39)	+10	0	+4	+5	−2	+7
35 nonchalant (27)	+10	+3	+2	+5	+4	+10
36 sad (41)	+10	+2	0	+2	0	+3
37 bored (80)	+8	+6	+7	+6	+4	+12
38 antagonistic (46)	+7	+3	+1	+6		+1
39 downhearted (48)	+7	+3	0	+3	+4	+5
40 hostile (96)	+7	+5	+2	+1	+1	+3
41 skeptical (116)	+7	0	−2	−2	0	+7
42 unhappy (61)	+7	+3	+2	+2	−1	+4
43 defiant (85)	+5	0	−1	−1	−4	+4
44 nervous (82)	+5	−2	−2	0	0	+4
45 remorseful (72)	+5	+3	0	+1	−4	−1
46 upset (45)	+5	+1	−2	−2		+4
47 annoyed (64)	+4	+3	+1	+4	−1	+3
48 suspicious (77)	+4	+5	+2	0	−1	+3

TABLE 1 (*continued*)

Adjective ( ) Ref. No., Table 2	100 mg 1955	80 mg 1958	80 mg '59 I	80 mg '59 II	50 mg 1954	100 mg 4 hr., '55
49 timid (54)	+4	−2	+1	+1	−1	0
50 blue (100)	+3	−3	−1	+2	0	+4
51 changeable (37)	+3	+5	0	+5	+5	−4
52 dissatisfied (81)	+3	+1	−4	+4	−10	+1
53 jittery (32)	+3	+3	+2	+2	+1	0
54 rebellious (97)	+3	0	−5	−1	0	+3
55 reckless (69)	+3	−1	0	−2	0	+4
56 self-contained (44)	+3	+2	0	+7	−1	−1
57 shy (53)	+3	0	+2	+2	−1	+4
58 angry (102)	+1	−1	−1	0	−1	−3
59 ashamed (63)	+1	+1	0	+2		+1
60 brooding (94)	+1	+2	+1	+7	−2	+8
61 guilty (86)	+1	0	0	0		0
62 sarcastic (62)	+1	−3	−2	+2	−1	−1
63 relaxed (121)	0	+3	+8	+7	+4	−1
64 self-conscious (87)	0	+2	+5	+1	−4	−1
65 uninhibited (33)	0	+1	−2	−5	+1	−3
66 cautious (107)	−1	−1	+3	+1	+2	−5
67 impatient (60)	−1	+10	+1	−2	−5	−3
68 keyed-up (40)	−1	+1	−2	−3	+1	0
69 peaceful (11)	−1	−7	+4	+5	+11	+4
70 apprehensive (36)	−3	+1	−2	0	−10	0
71 easygoing (75)	−3	−10	+1	−5	+8	+4
72 emotional (122)	−3	0	+1	−1	0	+1
73 sexy (22)	−3	−3	0	0	0	−3
74 dominating (56)	−4	−2	−2	−6	0	−3
75 original (14)	−4	−3	−3	−1	+6	−4
76 patient (115)	−4	−4	−2	+2	+6	−1
77 willful (78)	−4	−2	−5	−6	+1	−1
78 egotistic (52)	−5	−4	0	−2	−4	+1
79 trustful (92)	−5	−8	−4	−4	+2	−10
80 amused (93)	−7	−7	−8	−1	−1	−7
81 anxious (112)	−7	−5	−5	−2	−6	0
82 fearless (88)	−7	−3	−3	−2	−4	−7
83 forceful (20)	−7	+1	−4	−8	−4	−1
84 kindly (124)	−7	−7	−6	−5	+2	−7
85 boastful (71)	−8	−4	0	−2	−2	−7
86 carefree (84)	−10	−4	−2	−4	0	−7
87 elated (67)	−10	−3	−1	−2	−4	−10
88 friendly (111)	−10	−16	−5	−10	−2	−8
89 optimistic (126)	−10	−10	−11	−13	0	−11
90 warmhearted (109)	−10	−11	−4	−2	0	−10
91 witty (55)	−10	−4	−2	−2	+1	−14
92 excited (58)	−11	−6	−2	−4	−4	−4
93 humorous (104)	−11	−4	−7	−3	−1	−8
94 mischievous (24)	−11	−2	0	−4	−6	−7
95 aggressive (21)	−12	−6	−12	−12	−9	−8
96 engrossed (12)	−12	−4	−2	−6	0	−7
97 generous (95)	−12	−6	−2	−2	−2	−7
98 obliging (89)	−12	−11	−4	−7	+5	−11
99 work-oriented (70)	−12	−11	−12	−16	−4	−11

TABLE 1 (*continued*)

Adjective ( ) Ref. No., Table 2	100 mg 1955	80 mg 1958	80 mg '59 I	80 mg '59 II	50 mg 1954	100 mg 4 hr., '55
100 enterprising (38)	−14	−6	−12	−8	−4	−8
101 joyous (120)	−14	−2	+1	−2	+2	−10
102 meticulous (105)	−14	−8	−1	−2	+2	−10
103 resourceful (76)	−14	−6	−5	−6	+5	−8
104 careful (117)	−15	−6	0	−7	−1	−5
105 masterful (108)	−15	−4	−3	−5		−12
106 task-involved (42)	−15	−8	−8	−14	−6	−5
107 genial (98)	−16	−5	−6	−8	−1	−11
108 assertive (15)	−18	−10	−6	−10	−10	−10
109 cooperative (114)	−18	−12	−7	−12	0	−7
110 goodnatured (125)	−18	−10	−10	−12	−1	−10
111 satisfied (17)	−18	−4	−3	−8	+2	−14
112 confident (113)	−19	−11	−8	−3	+1	−14
113 decisive (118)	−19	−12	−8	−7	−5	−11
114 self-confident (99)	−19	−6	−10	−10	+1	−5
115 attentive (74)	−20	−6	−12	−12	−2	−8
116 hardworking (59)	−20	−8	−10	−14	−4	−11
117 industrious (119)	−20	−12	−11	−10	−9	−10
118 happy (128)	−22	−11	−9	−12	0	−12
119 talkative (106)	−22	−10	−2	−6	−1	−12
120 cheerful (129)	−23	−15	−13	−13	−8	−19
121 enthusiastic (68)	−23	−10	−6	−9	−10	−16
122 effective (51)	−24	−12	−8	−15	−1	−16
123 vigorous (73)	−24	−5	−8	−12	−6	−19
124 sociable (101)	−26	−11	−4	−14	+2	−22
125 eager (43)	−27	−12	−12	−13	−8	−20
126 ambitious (130)	−29	−23	−13	−18	−1	−20
127 efficient (31)	−29	−11	−8	−11	+1	−16
128 energetic (123)	−29	−8	−11	−21	−12	−19
129 business-like (110)	−33	−11	−12	−17	−9	−20
130 active (35)	−34	−15	−22	−28	−11	−26
Mean change	14.1	6.5	6.2	7.8	4.4	9.3

ignored in evaluating the much larger changes after dramamine.[4] Whichever course is chosen, subtracting or not subtracting, one commits a certain amount of error. Changes after placebo are in part random effects (which cannot

---

[4] The reader may wish to perform these subtractions. The design was such that certain subtractions are the only appropriate ones. In each case the *total change scores* for individual adjectives in Table 2 are to be subtracted from those in Table 1. The cross references for matching adjectives are in the left-hand column of Table 1. Subtract as follows:

Table 1	*Table 2*
1955 column	Mean of 1955 I and 1955 II columns
1958 column	Mean of 1958 I, 1958 II and 1958 III columns
1959 I column	1959 I column
1959 II column	1959 II column
1954 column	1954 column
1955 (4 hr. column)	1955 (4 hr. column)

properly be subtracted) and in part systematic (which can be subtracted in most cases).

The numbers in Table 1 are *total change scores* for individual adjectives for the *S*s in each experimental administration of dramamine. The first number in Table 1 is +50. This means that in the 1955 experiment the total change for all *S*s from before drug to during drug on the adjective #1, drowsy, was fifty adjusted score steps toward being *more* drowsy. (The theoretical maximum change score would be +60, if all *S*s did not check "drowsy" before drug and double-checked it after drug). The actual unadjusted total change score for 22 *S*s was +37. One can read across and see that in the five other assessments of effects of dramamine on the adjective drowsy, the adjusted change scores were +29, +27, +34, +26 and +34.

The adjectives have been rearranged from the original alphabetic list. They appear in rank order of the magnitude of the effect of the 100 mg administration and are listed from greatest increases, i.e., drowsy, sluggish, weary, etc., through no change, i.e., relaxed, self-conscious and uninhibited, to the greatest decrease, i.e., active, which is —34. Empty cells in the 1954 data are for adjectives not in common between the two ACLs. The columns (except for the last) have been arranged in descending order of dose size. The last column is for a four-hour-after-drug assessment, obtained only in 1955.

2. *Effects of Dramamine*

An adjective change score preceded by a plus sign indicates that the *S*s felt more that way after drug; by a minus sign, less that way. Inspection of the table is needed to see the nature of these changes. In general we will show that dramamine had sedative effects, although it would be more accurate to describe the effects of small doses (50 mg) as being tranquilizing. The interested reader who wishes to try to verify this conclusion for himself is urged to classify the adjectives into groups for similarity, match them with their approximate antonyms, and then explore the emotional and motivational areas affected by the experimental procedures.

We have empirically found that an N of 30 yielding a change score of seven or larger for an adjective gives a meaningful change. Consideration of adjectives with even smaller change scores is recommended in that many continue to have internal consistency and fall into reasonable groups of emotional similarity.

a. Effects of various doses. The nature of the changes two hours after 100 mg of dramamine can readily be seen by reading the list of adjectives, since these have been arranged in order of size of change score for that dose. The *S*s reported themselves as *more* drowsy, sluggish, languid, withdrawn, dizzy, silent, depressed and antagonistic, and as *less* energetic, business-like, effective, talkative, decisive, confident, good-natured, witty and carefree.

The effects of 80 mg and of 100 mg were qualitatively similar except that after 100 mg such emotionally negative adjectives as #25 disturbed, #29 belligerent, #40 hostile, and #38 antagonistic were more often checked. We feel that this was, at least in part, a derived or secondary effect (13, pp. 199-200) in that the *S*s were disturbed and antagonistic *because* we had made them unpleasantly drowsy and inefficient, not simply because the drug produced antagonism.

After 50 mg of dramamine a number of adjectives which one would class otherwise than in a sedation category were unchanged or moved toward the positive mood direction. By comparison of the 50 mg column with the columns for larger doses and with the changes after placebo, adjectives were identified which (*a*) indicated an improvement of mood or no change (adjusted for placebo change) from 50 mg and (*b*) were at least seven score units different from the changes with larger doses. The results are as follows:

After 50 mg (as compared to 80 or 100 mg) the *S*s were *less* #8 dull, #15 withdrawn, #16 dizzy, #20 confused, #22 doubtful, #27 disinterested, #52 dissatisfied, #70 apprehensive, and #23 indifferent.

They were *more* #69 peaceful, #71 easygoing, #75 original, #76 patient, #84 kindly, #88 friendly, #89 optimistic, #98 obliging, #99 work-oriented, #103 resourceful, #109 cooperative, #110 good-natured, #112 confident, #114 self-confident, #115 attentive, #118 happy, #124 sociable, #126 ambitious, and #127 efficient.[5]

It seemed worthwhile to adjust for placebo changes when dealing with the dose of 50 mg, since these placebo data showed the operation of systematic factors. We have done this for all adjectives and find that 19 of them are shifted from change scores of less than seven to scores of seven or more, while six shifted to non-meaningful levels. These are shown in Table 1a. The results are concordant with those of the preceding analysis. It may also be noted that most of the adjectives in Table 1a fall into clusters of related mood states, thus lending credence to the propriety of adjusting for placebo changes in this instance.

---

[5] The 1954 data on 50 mg were obtained from a 200 word ACL containing 76 adjectives not duplicated in the 130 word list. These adjectives were removed from the list because some had been insensitive. The data of these 76 adjectives show that the *S*s reported themselves as being relatively *more* calm, capable, contented, gentle, submissive, reserved, and aloof, and as being relatively *less* tense, frustrated, erratic, and earnest than they were under placebo. These effects are concordant with those for the 124 adjectives in common with the 130 word ACL.

A rough quantitative comparison of the over-all effects of the three dose levels is given by the *mean adjective change scores* in each sampling (last entries at bottom of Table 1). There were differences in repetitions of the same dose (i.e., 80 mg data where mean change scores were 6.5, 6.1 and 7.8), and pronounced differences between doses (100 mg = 14.1, 80 mg = mean of 6.8, 50 mg = 4.7). A more satisfactory measure of these differences is by adjective cluster scores (13, p. 209), to be reported at a later time.

TABLE 1a
ADJECTIVES AFFECTED BY ADJUSTING CHANGE SCORES
50 mg Data

Adjective	50 mg	Placebo	Adjusted	Adjective	50 mg	Placebo	Adjusted
relaxed	+4	− 8	+12	tame	+ 6	− 1	+ 7
cooperative	0	−12	+12				
friendly	−2	−14	+12	lackadaisical	+16	+10	+ 6
sociable	+2	−10	+12	weary	+ 9	+ 5	+ 4
obliging	+5	− 6	+11	silent	+ 8	+ 5	+ 3
confident	+1	− 8	+ 9	cheerful	− 8	−10	+ 2
warmhearted	0	− 8	+ 8	energetic	−12	− 8	− 4
carefree	0	− 8	+ 8	business-like	− 9	− 5	− 4
ambitious	−1	− 9	+ 8				
engrossed	0	− 8	+ 8	forceful	− 4	+ 4	− 8
work-oriented	−4	−11	+ 7	impatient	− 5	+ 4	− 9
resourceful	+5	− 2	+ 7	confused	− 1	+10	−11
generous	−2	− 9	+ 7	mischievous	− 6	+ 9	−15
kindly	+2	− 5	+ 7				

b. Comparison of two and four hours after dose. In 1955 the *S*s returned to the laboratory for a third measure 4 hours after taking 100 mg of dramamine. These data are shown in the last column of Table 1. One can obtain a rough quantitative estimate of the relative effect two hours and four hours after taking the drug by noticing that the average change scores per adjective were 14.1 and 9.3. Since other information indicates that two hours is already past the peak effect, it can be seen that four hours is well past the maximum effect.

All of these *S*s completed a brief questionnaire in the evening. This indicated that they were no longer able to notice effects seven hours after the dose. (An ACL at this time might have detected residual effects or after-effects.)

c. Reports of drowsy and sleepy. Our data indicate more sedation than reported by others (3, 4). Because the dramamine literature mentions *S*s' reports of drowsiness as being infrequent, we determined each individual *S*'s two-hour change scores for drowsy and sleepy so as to be able to express effects in terms of their frequency of occurrence. The per cent of *S*s who increased on drowsy or on sleepy or on both was 86 per cent after 100 mg, 85 per cent after 80 mg and 67 per cent after 50 mg. If these are adjusted by subtracting the placebo changes in the same experiments: 18 per cent, 21 per cent and 21 per cent respectively, one obtains frequencies of 68 per cent, 64 per cent and 46 per cent.[6]

d. Consistency in consecutive experiments. If all doses had been the same (e.g., 80 mg of which there were three administrations), consistency could be rapidly judged by the consistency of plus or minus signs across each row, and, more finely judged by the variation in size of the change scores. Because dose level was related to kind of emotional change (especially in comparing 100 mg with 50 mg) consistency within rows is somewhat reduced (e.g., adjectives #25 disturbed, #29 belligerent, #118 happy, #124 sociable, #127 efficient). On the other hand, the over-all consistency from experiment to experiment and even from one dose level to another seems to us to have been very satisfactory for this type of work. It will be noticed that where a plus change (*S*s feel more that way) of seven or more appears for an adjective in the 100 mg dose column, it usually appears in the other columns. In the same way, where *S*s decreased their acceptance of an adjective (minus signs) by seven or more after 100 mg, other *S*s decreased acceptance under other doses.

Consistency was by no means perfect. This can be seen by inspecting the data of the three determinations of the effects of 80 mg. There are 19 adjectives in the 80 mg data for which one experiment deviated from the mean of the other two by seven or more. Languid, #9, is the first of these: the 1958 result of a +4 change score deviated by 15 from the mean of the two 1959 determinations.

When one inspects all 19 such deviant adjectives, it becomes evident that these were not usually chance deviations for individual adjectives, but, instead, that a number of closely related adjectives were affected. For example, in the 1958 data on 80 mg the adjectives differing most from the other two sets of data were #130 active, #83 forceful, #67 impatient (which were relatively more acceptable to the *S*s) and #9 languid, #12 lackadaisical, #11 quiet, #14 tame, #69 peaceful, #90 warm-hearted, #88 friendly (which were relatively less acceptable). In the 1959 I data, #71 easygoing and #124 sociable were more accepted; the adjective #31 dubious was less accepted. In the 1959 II data, #10 weak, #24 subdued, #26 dreamy, #28 close-mouthed and #19 silent were more accepted; #128 energetic was less accepted. This suggests that the 1958 data differed systematically from the 1959 data, but that deviant systematic factors within the 1959 data were not strong. The number of such deviant adjectives, 10, 3, and 6 in the three sets of data was not large.[7]

---

[6] One *S* (out of 112) had change scores in opposite directions on drowsy and sleepy. Such occasional embarrassments can be attributed to *S*'s carelessness, or (more likely) to *S*'s atypical self-instruction about the interpretation of words. The latter is exemplified by four *S*s in one year's data who increased on both drowsy and sleepy, yet changed from a double check on tired to leaving it blank. We are inclined to believe that the word tired was used by these *S*s when entitled to the feeling after their morning work, and not after luncheon leisure when it became inappropriate while drowsy and sleepy became accurate.

A simple way of assessing the extent to which one experiment with dramamine gives the same results as another is to compute the correlation of the 130 change scores in one experiment with the 130 change scores in another. This gives one a direct assessment of the repeatability of these experiments. These correlations have been computed for each experiment where 80 mg of dramamine was administered. The correlations are:

$$r_{59I\cdot 59II} = .92$$

$$r_{58\cdot 59I} = .84$$

$$r_{58\cdot 59II} = .86$$

$$r_{58\cdot}\ \frac{59I + II}{2} = .87$$

e. Internal consistency of emotional changes after drug. One of our criteria for acceptance or non-acceptance of a finding is the internal consistency of the data (13, pp. 179-180). Inspection of Table 1 shows that this was present to a satisfactory degree. Thus the 100 mg data in the first column show both that synonyms were similarly affected by the drug in most cases (a kind of reliability check) and that closely related states were affected in the same direction and to roughly the same degree (a kind of construct validity check), e.g., #1 drowsy and #5 sleepy; #6 lazy, #9 languid and #12 lackadaisical; #3 weary, #4 tired and #7 washed-out; #11 quiet, #14 tame and #19 silent; #20 confused, #22 doubtful and #25 disturbed; #13 drifting and #26 dreamy; #29 belligerent, #38 antagonistic and #40 hostile. Antonyms of the preceding adjectives were moved in the opposite direction, e.g., #123 vigorous, #128 energetic, #130 active. Section *a, c* and *d* above give other examples of internal consistency in various parts of Table 1.

---

[7] Search for the reasons for these differences is beyond the scope of the present paper. They need to be sought not only in special factors affecting the post-drug report, but also in factors in the pre-drug report of each *S*. These factors may be such things as weather, campus events, temporary attitudinal changes in *S*s such as lessened interest, etc.

These changes under dramamine were not simply a shift of *all* related emotional states in the same direction, but were to some degree specific to adjective clusters. If *all* were equally affected by drug then tiredness, hostility, depression and anxiety (all of which are moderately to highly correlated in the pre-drug data) would be expected to be affected together. They were not all affected together at any dose level in the range we used. The administration of drug separated certain emotional clusters which ordinarily are positively correlated.

3. *Change Scores After Placebo*

Table 2 displays the change scores of placebo sessions in much the same way as the dramamine data. The order of adjectives is by over-all magnitude of change scores (average shown in column 1) for all two-hour placebo sessions in all experiments, totalling 230 subject sessions. Cross reference to the dramamine data is given by the serial number in parenthesis. As in Table 1, each entry has been adjusted to N = 30 and decimals rounded off. The four-hour placebo data for 1955 (last column) are the means of the two four-hour placebo sessions. Such four-hour data are slightly more variable than two-hour data.

If we judge results by combining the eight two-hour placebo administrations (column 2), the change scores of none of the 130 adjectives were large enough to be of much consequence. The five largest mean change scores were +4.6 lackadaisical, —4.4 ambitious, —4.0 cheerful, —3.2 happy, and +3.1 sluggish. Of the 130 adjectives, 125 had average change scores of less than 3.0, 109 of less than 2.0 and 66 of less than 1.0. (After 100 mg of dramamine there were only 12 adjectives with change scores of less than 3.0.) This seems to be adequate evidence (for the over-all data) of success in reducing placebo effects.

Judging results of individual determinations of two-hour placebo changes, one notes that there are change scores of seven or more in each column of data except one. The larger change scores are found in the 1954, 1955 I, 1955 II and 1959 II data; the data of 1958 I, II and III and of 1959 I are mostly without changes after placebo. Individual adjectives in individual years showed a total of 63 occasions (out of a possible 1034) on which the change score was seven or larger. Tabulation of the adjectives, done separately for each set of data, suggests to us that the observed changes were not due to placebo "effects" to any significant degree. On the contrary, they mostly appear to have clustered in such a way as to point to changes in mood at the time of the

TABLE 2

EFFECTS OF PLACEBO ON MOODS, EMOTIONS AND MOTIVATIONS. TOTAL *Change Scores* FOR EACH GROUP OF SS
(Scores adjusted to N = 30)

Adjective ( ) Ref No., Table 1	Av. of all Experiments	1954	1955 I	1955 II	1958 I	1958 II	1958 III	1959 I	1959 II	4 hr., 1955 Av. of I & II
1 lackadaisical (12)	+4.6	+10	—1	+12	+3	+5	+4	+2	+2	+3.5
2 sluggish (2)	+3.1	+1	+1	+10	+3	+5	0	+1	+4	+2.5
3 sleepy (5)	+2.8	0	+4	+7	+7	+1	+3	0	0	+2.0
4 subdued (24)	+2.8	+2	+5	+5	—1	0	+6	+2	+3	+1.5
5 indifferent (23)	+2.5	+5	+1	+5	—1	+1	+2	+2	+5	0
6 tired (4)	+2.5	+4	+4	+10	+4	—2	—2	0	+2	+7.5
7 lazy (6)	+2.0	+4	+1	+7	0	+2	—1	+1	+2	+3.5
8 drifting (13)	+1.9	+4	—1	+4	—4	+3	+5	+2	+2	+2.5
9 dreamy (26)	+1.6	—5	+1	+7	+3	+4	+2	+1	0	+2.5
10 leisurely (17)	+1.6	+2	0	+7	+1	+3	—2	+2	0	+5.0
11 peaceful (69)	+1.6	+4	0	+7	—2	—1	+1	+1	+3	+8.0
12 engrossed (96)	+1.5	—8	+5	+3	+3	+6	+3	—4	+4	+2.0
13 languid (9)	+1.5	+4	+3	—1	+1	+1	+2	+2	0	—1.0
14 original (75)	+1.5	+1	+4	+3	+1	+1	—1	+3	0	+3.0
15 assertive (108)	+1.4	+1	+4	+1	0	+5	—1	+3	—2	—6.5
16 light-headed (18)	+1.4	+1	+1	+5	—2	+1	+2	+2	+1	+4.0
17 satisfied (111)	+1.4	+4	+4	+4	+2	+3	—3	+3	—6	+5.0
18 withdrawn (15)	+1.4	+6	+3	—1	+2	—3	0	+2	+2	—2.5
19 headache (32)	+1.2	+5	+4	0	—1	+1	—2	+1	+2	+3.5
20 forceful (83)	+1.2	+4	+10	0	+1	0	—1	—3	—1	+2.5
21 aggressive (95)	+1.1	—1	—5	—5	+7	+5	+3	0	+5	—7.0
22 sexy (73)	+0.9	+1	+1	—1	+1	+2	+1	+2	0	+4.0
23 confused (20)	+0.9	+10	—5	—5	+2	0	+4	+2	—1	—6.5
24 mischievous (94)	+0.8	+9	—1	0	+2	+1	—2	—2	—1	—2.0
25 silent (19)	+0.8	+5	0	0	+1	—2	+2	+2	—2	—2.0
26 depressed (30)	+0.6	0	+3	+1	—1	—2	+1	+1	+2	+1.5
27 nonchalant (35)	+0.6	—1	0	+4	—1	0	+2	0	+1	+4.5
28 weak (10)	+0.6	0	+3	—3	0	0	+1	+5	—1	+3.5
29 disinterested (27)	+0.5	+4	—5	+1	—1	+2	+1	+2	0	—2.0
30 drowsy (1)	+0.5	+4	—3	+8	+2	—3	—2	—3	+1	0

TABLE 2 (*continued*)

Adjective ( ) Ref No., Table 1	Av. of all Experiments	1954	1955 I	1955 II	1958 I	1958 II	1958 III	1959 I	1959 II	4 hr., 1955 Av. of I & II
31 efficient (127)	+0.5	— 1	+ 8	0	0	+2	—1	+2	—6	+5.0
32 jittery (53)	+0.5	+ 2	+ 7	— 1	+1	—3	—1	—3	+2	—2.5
33 uninhibited (65)	+0.5	— 4	+ 1	+ 3	—1	+1	+1	+1	+2	—0.5
34 weary (3)	+0.5	+ 5	+ 1	0	+2	—6	+1	0	+1	+0.5
35 active (130)	+0.4	— 2	— 8	+ 1	+4	+6	+1	+3	—2	—4.0
36 apprehensive (70)	+0.4	— 2	0	+ 1	+6	—2	+3	—1	—2	—4.0
37 changeable (51)	+0.4	— 1	+ 3	+ 4	—1	—1	+1	+1	—3	—7.0
38 enterprising (100)	+0.2	0	+ 1	— 5	0	+6	+1	0	—1	—3.5
39 intoxicated (34)	+0.2	— 1	+ 1	+ 3	0	0	—2	+1	0	0
40 keyed-up (68)	+0.2	+ 4	0	+ 1	+5	—3	+1	—5	—1	—3.0
41 sad (36)	+0.2	+ 1	0	— 1	+4	—3	+2	—1	0	—2.0
42 task-involved (106)	+0.2	— 8	+ 4	0	+4	+1	+1	+2	—2	—1.5
43 eager (125)	+0.1	+ 1	+ 7	+ 1	—1	+1	—2	—2	—4	0
44 self-contained (56)	+0.1	+ 4	— 3	+ 1	+1	—1	0	+1	—2	+1.0
45 upset (46)	+0.1		— 1	+ 3	+1	—1	+2	—2	—1	—1.5
46 antagonistic (38)	0		0	+ 1	—1	—1	—2	+2	+1	0
47 dizzy (16)	0	+ 1	— 1	— 1	—1	+3	—2	+1	0	—2.0
48 downhearted (39)	0	+ 1	+ 1	0	+2	—3	+1	—1	—1	—5.5
49 dull (8)	0	+ 6	— 3	— 3	+1	—1	—1	—1	+2	—3.0
50 close-mouthed (28)	—0.1	+ 1	0	0	—1	—3	—1	+1	+2	—0.5
51 effective (122)	—0.1	+ 1	+ 3	+ 1	—2	+1	+1	—3	—3	+1.5
52 egotistic (78)	—0.1	— 1	— 3	+ 3	+5	0	—2	—4	+1	—2.0
53 shy (57)	—0.1	— 2	— 3	+ 4	+1	0	0	+1	—2	—1.0
54 timid (49)	—0.1	+ 2	— 1	— 1	—1	0	0	+1	—1	—1.0
55 witty (91)	—0.1	+ 1	— 4	0	—2	+4	—1	—2	+3	—1.5
56 dominating (74)	—0.2	— 1	— 4	0	+1	—1	0	—1	+4	—0.5
57 doubtful (22)	—0.2	+ 1	— 3	— 1	+4	—5	+1	+1	0	—4.5
58 excited (92)	—0.2	— 1	— 1	— 1	0	+2	0	0	—1	—4.0
59 hardworking (116)	—0.2	— 1	0	— 4	+3	—2	0	+2	0	—1.0
60 impatient (67)	—0.2	+ 4	— 7	— 1	+3	—6	—2	+5	+2	—9.5
61 unhappy (42)	—0.2	0	+ 3	— 1	+1	—1	0	—2	—2	—2.0
62 sarcastic (62)	—0.2	— 1	0	0	—3	+1	0	+2	—1	—2.0
63 ashamed (59)	—0.3		— 1	— 1	+2	0	—1	0	—1	—1.5

TABLE 2 (*continued*)

Adjective ( ) Ref No., Table 1	Av. of all Experiments	1954	1955 I	1955 II	1958 I	1958 II	1958 III	1959 I	1959 II	4 hr., 1955 Av. of I & II
64 annoyed (47)	—0.4	+2	—3	—1	+1	—4	0	—1	+3	—0.5
65 detached (21)	—0.4	+1	—4	—4	0	+3	+4	—1	—2	+1.5
66 dubious (31)	—0.4	+1	0	—4	+2	—4	—1	+1	+2	+3.0
67 elated (87)	—0.4	—2	—4	+3	—1	0	—2	+1	+2	—3.0
68 enthusiastic (121)	—0.4	—2	+4	—4	+3	+5	+1	—2	—8	—4.0
69 reckless (55)	—0.4	—1	—1	+3	—1	0	—2	—3	+2	—0.5
70 work-oriented (99)	—0.4	—11	+1	+4	+4	+2	—2	+1	—2	—5.0
71 boastful (85)	—0.5	+1	—7	+4	+1	0	—3	0	0	—3.5
72 remorseful (45)	—0.5	—1	+1	0	+1	—3	—2	0	0	—1.0
73 vigorous (123)	—0.5	—4	+8	—8	+3	—3	+3	+3	—6	+0.5
74 attentive (115)	—0.6	—5	+1	+1	—1	+2	+1	—2	—2	—0.5
75 easygoing (71)	—0.6	0	—3	+1	—2	+3	—3	+4	—5	—2.5
76 resourceful (103)	—0.6	—2	+5	—4	+3	—1	—2	—1	—3	—2.5
77 suspicious (48)	—0.6	—1	—1	—5	—1	—2	+3	+2	0	—2.0
78 willful (77)	—0.6	—2	0	+1	—1	0	—2	+1	—2	0
79 disturbed (25)	—0.7		+3	0	—4	—2	—2	+1	—1	—4.0
80 bored (37)	—0.8	—1	—3	—1	0	—1	+1	0	—1	—1.5
81 dissatisfied (52)	—0.8	+2	+3	—3	—1	0	—1	—3	—3	—2.0
82 nervous (44)	—0.8	+1	0	—3	+2	—3	+1	—3	—1	—2.0
83 washed-out (7)	—0.8	0	+3	—5	—2	+1	—3	—2	+2	—1.5
84 carefree (86)	—0.9	—8	—3	+3	+1	+3	—3	0	0	+2.0
85 defiant (43)	—0.9	—2	—1	—1	0	—4	0	—1	+2	0
86 guilty (61)	—0.9		+3	—4	—2	—1	0	—2	0	—2.5
87 self-conscious (64)	—0.9	+2	—5	0	0	—4	0	0	0	—4.0
88 fearless (82)	—1.0	+1	—1	—1	—1	0	—3	—1	—2	0
89 obliging (98)	—1.0	—6	—3	+1	+6	0	—4	0	—2	0
90 quiet (11)	—1.0	0	—5	—1	—3	—2	—3	+5	+1	—0.5
91 tame (14)	—1.0	—1	+1	—5	—4	0	—1	+2	0	—2.0
92 trustful (79)	—1.0	—4	—3	0	0	+4	—3	0	—2	—5.5
93 amused (80)	—1.1	—6	—3	+4	—1	0	—2	+2	—3	+2.5
94 brooding (60)	—1.1	—1	+1	—4	0	0	—5	—1	+1	—3.0
95 generous (97)	—1.1	—9	+1	—3	+4	0	—3	+5	—4	—1.0
96 hostile (40)	—1.1	—2	—1	0	—1	—3	0	—2	0	0

TABLE 2 (*continued*)

Adjective ( ) Ref No., Table 1	Av. of all Experiments	1954	1955 I	1955 II	1958 I	1958 II	1958 III	1959 I	1959 II	4 hr., 1955 Av. of I & II
97 rebellious (54)	—1.1	— 1	0	— 1	+2	—3	—2	—2	— 2	—1.5
98 genial (107)	—1.2	+ 1	— 5	— 1	+2	+1	—4	+4	— 8	—2.0
99 self-confident (114)	—1.2	0	+ 1	— 5	—3	+2	—1	+3	— 7	—3.0
100 blue (50)	—1.4	— 6	0	+ 1	—1	—2	—1	—1	— 1	+0.5
101 sociable (124)	—1.4	—10	0	0	—3	+3	—4	+5	— 2	—2.0
102 angry (58)	—1.5	+ 1	— 5	— 1	0	—5	—1	—1	0	—2.0
103 belligerent (29)	—1.5	— 4	— 3	— 3	—1	—1	0	0	0	—2.0
104 humorous (93)	—1.5	— 1	— 3	— 7	—2	+5	—2	0	— 2	—5.5
105 meticulous (102)	—1.5	+ 1	+ 1	—11	—2	+2	+2	0	— 5	—0.5
106 talkative (119)	—1.5	+ 2	— 8	— 1	—1	+1	—7	—1	+ 3	—0
107 cautious (66)	—1.6	+ 1	— 1	— 7	—6	+1	+3	—6	+ 2	—3.0
108 masterful (105)	—1.6		— 1	— 5	—1	0	0	+2	— 6	—0.5
109 warm-hearted (90)	—1.6	— 8	0	+ 4	—3	—3	—3	+1	— 1	—1.0
110 business-like (129)	—1.8	— 5	+ 4	— 5	—9	0	+3	+4	— 6	—1.5
111 friendly (88)	—1.8	—14	— 7	+ 5	0	—2	+2	+4	— 2	—0.5
112 anxious (81)	—1.9	— 4	— 8	+ 1	—5	—2	+5	+1	— 3	—2.0
113 confident (112)	—1.9	— 8	+ 3	0	—1	+1	—1	—7	— 2	+2.0
114 cooperative (109)	—1.9	—12	0	+ 3	—2	+3	+2	—1	— 8	+2.0
115 patient (76)	—1.9	+ 1	— 3	— 4	—1	0	0	—7	— 1	+3.0
116 skeptical (41)	—1.9	0	— 5	— 8	+2	—1	—1	—1	— 1	—3.0
117 careful (104)	—2.0	+ 2	0	—10	—8	+5	+1	—2	— 4	—2.0
118 decisive (113)	—2.0	+ 1	— 4	+ 1	—4	—2	—3	—3	— 2	+0.5
119 industrious (117)	—2.0	— 2	+ 5	— 8	+1	0	—4	—1	— 7	—2.5
120 joyous (101)	—2.1	— 2	—16	+ 1	—1	+1	—2	+1	+ 1	—6.5
121 relaxed (63)	—2.1	— 8	— 5	+ 1	—3	+2	—2	0	— 2	+3.5
122 emotional (72)	—2.2	— 4	— 5	— 4	—1	—2	—1	0	— 1	—4.0
123 energetic (128)	—2.2	— 8	— 4	— 1	+2	0	—5	0	— 2	—1.5
124 kindly (84)	—2.5	— 5	— 5	— 3	—2	+1	—2	+1	— 5	—3.0
125 goodnatured (110)	—2.6	— 1	— 4	0	—3	+4	—5	—4	— 8	+1.5
126 optimistic (89)	—2.6	— 5	— 4	0	+3	+1	—2	—2	—12	—3.5
127 impulsive (33)	—2.8	— 2	— 5	— 4	—2	—1	—4	—2	— 2	—4.5
128 happy (118)	—3.2	— 1	— 7	— 5	0	—1	—4	0	— 8	+0.5
129 cheerful (120)	—4.0	—10	— 3	— 5	—3	—5	+1	+3	—10	—2.0
130 ambitious (126)	—4.4	— 9	— 4	— 4	—3	—2	—7	—2	— 4	—5.0
Mean adjective change score		3.2	2.9	2.9	2.0	2.0	1.9	1.8	2.4	

afternoon sessions which were not produced by attitudes toward the capsules. This was a depression of mood in 1954, 1955 II and 1959 II. In 1955 I, there was no consistent pattern. We believe that in 1954 and 1955 II we did not achieve enough afternoon interest in the *S*s to overcome the slight after-lunch depression of mood. In 1955 I, when mood changes were inconsistent after lunch, our experiment ran afoul of a series of campus social events during which about half of the *S*s completed their placebo I session. In 1959 II a noticeable slight uncooperativeness developed over a campus rumor that we were giving "hate pills" (13, p. 198). This should be called a true placebo "effect."

The numbers of change scores of seven or above for each column in Table 2 were respectively 17, 13, 16, 4, 0, 2, 2, and 9. The ascending order of these from 0 to 17 roughly corresponds to our judgment of the goodness of control of the attitudinal and personal variables of the *S*s in each set of data.

Another way of estimating the stability of placebo data is to calculate the mean adjective change score, shown at the bottom of Table 2. This shows our most stable placebo data to be those for 1958 I, II and III and for 1959 I. The mean adjective change scores of 1.8 to 2.0 score points then obtained appear to us to be rather remarkable stability for an N of 30.

Yet another approach to the question of whether systematic factors were acting is to compute the correlation of trials for the 130 pairs of adjective change scores. The correlations are as follows:

$r_{55I \cdot 55II} = +.05$

$r_{58I \cdot 58II} = -.07$

$r_{58I \cdot 58III} = -.02$

$r_{58II \cdot 58III} = +.04$

$r_{59I \cdot 59II} = -.07$

The very small size of these correlations indicates that systematic factors were not operating.

It is our belief that *S*s who are kept interested, cooperative and motivated can perform the introspective task required by the ACL while maintaining stability of mood and attitude. In other words, we believe that we have demonstrated that placebo "effects" can be successfully minimized in male college student *S*s.

D. Discussion

1. *Placebo Data*

One purpose of this paper was to present certain of our placebo data. In previous papers we have displayed similar data (1, 5, 6, 10) and have described procedures designed to minimize so-called placebo "effects" (13). The present paper has offered evidence that data can be obtained which show little or no effect attributable to expectations or attitudes triggered by the taking of the capsules. One virtue of such an accomplishment is that the true effect of a drug can then be found, within the limits set by the sensitivity, validity and other aspects of the measurement procedures. A second virtue is that such placebo data can also be used as normative data to assess the effects of a host of personal and environmental variables on the moods, emotions and motivations included in the ACL, as well as for studies of interrelations of emotional states under normal conditions. *Such studies are better done in conjunction with drug studies rather than separately, since the Ss' search for drug effects provides superior interest and motivation for accurate introspection.*

2. *Use of Double Controls in Psychopharmacology*

The need for double controls by use of both placebo and dramamine is well-illustrated by our study of SC-2491 which turned out to be an inactive drug. Had we used only placebo and obtained a near total absence of changes after placebo as well as after SC-2491, such a result could be attributed to insensitive methods. We would then have had no basis for asserting that the drug was inactive. On the other hand, the inclusion of dramamine and its demonstration of sensitive methods, permitted the conclusion that the drug was inactive. Furthermore, if one studied benzedrine in one experiment and dexedrine in another, one could not evaluate quantitative differences between them unless one had data on a standard drug to indicate the sensitivity of the method in each experiment. In one experiment we included as many as six known drugs, thus further contributing to a test of sensitivity and validity. It is proof of validity insofar as the results correspond with the generally accepted effects of the known drugs. Our earlier paper has presented other aspects of this problem (13, p. 180).

If one is to use a standard drug, it must be shown that the drug gives comparable results from one experiment to another. We recognize that our six assessments of the effects of dramamine are complicated by the variables of dose size, time after dose (four hour data of 1955) and learning by repetition (1959 II data). In spite of these variables, the general impression is that the data are consistent enough to make dramamine a good indicator of whether the measures are sensitive and valid. Even when not entirely consistent the locations of the inconsistencies shed light on unusual factors affecting the experiment (see p. 267, above). We have done thirteen other similar experiments with dramamine, mostly unreported, which gave concordant results (5, 6, 11). Finally, the literature (e.g., 3, 4, 7) in general supports our findings.

3. *The Nature of the Dramamine Effect*

When a drug gets identified in the public mind with a particular classification such as *motion sickness preventive* or *tranquilizer*, it may be difficult to get consideration of other possible uses. Thus, while dramamine has been recognized as an anti-emetic and especially as a motion sickness preventive, its sedative and tranquilizing properties are not so widely known. These four experiments seem to leave no room for doubt of such effects when it is administered to male college students.

Dramamine administered to patients or travellers does not regularly elicit reports of sedation, yet most of our *S*s gave such reports. This difference should not be construed as impugning either clinical reports or our laboratory data. The air traveller susceptible to motion sickness may correctly report an elevation of mood produced by being simultaneously tranquilized and free from motion sickness. Furthermore, while patients may complain about a drug (or be very pleased) only a few of them adopt a critical, sensitive introspective attitude (12). Their attention is too much focused on their symptoms, or lack of them, to report side effects. By contrast, our *S*s are unbiased, and their use of the ACL plus the careful experimental environment encourage them to become introspective.

4. *The Adjective Check List as a Measuring Device*

The empirical findings from use of the ACL as a method for assessing drug effects have enough face validity and repeatability to justify its use. On the other hand, the ACL falls short of meeting some of the criteria of good measurement methods as previously set forth by us (9, pp. 52-61).

Some of the ways in which the ACL falls short of perfection are the following: the scoring scale is probably not linear, differing between both adjectives and *S*s. Some adjectives are stronger than others, e.g., angry *vs* irritable. The frequency of checking varies for different adjectives so that the possibilities of changes are different. For example, in these experiments "lazy" was left blank before drug by 96 per cent of *S*s while "active" was left blank by only 14 per cent and double-checked by 44 per cent; hence the former had little latitude for decreases in checking whereas the latter had considerable latitude. Adjectives may have slightly different meanings for different *S*s and *S*s may adopt different self-instructions about checking. The same emotional state may cause some *S*s to check different adjectives than do others, even though both are checking in the same emotional cluster. Drugs may affect the readiness to check, for example, dramamine decreased it in some earlier work. Certain drugs, e.g., secobarbital, also affect the *S*s' introspective insights. Finally, some emotional states are less accurately reported than others, often because of less awareness of them in some *S*s.

The ACL, as used by us for assessment of drug effects in male college students, has the following demonstrable advantages:

a. It is reliable. Instead of a test of the reliability of the instrument, we have a check of the reliability of the entire procedure, within and between experiments. The change scores obtained for the 130 adjectives in one administration of dramamine correlated to a satisfactory degree with those obtained in a second administration ($r = +.92$). The correlation between 1958 dramamine change scores and 1959 dramamine change scores for the 130 adjectives was $r = +.87$.

b. It is valid. The adjective changes after dramamine are consistent with the literature and in many important particulars verified by objective data (largely unpublished). Our unpublished work on amphetamines, caffein and others show that the ACL results correspond to the literature in most important aspects.

*c. It is indefinitely repeatable in well-motivated S*s. The correlation of $r = +.92$, mentioned above, was between two 1959 assessments with an average of 14 ACLs intervening. The 1954 and 1955 data were obtained from *S*s who had completed 50 to 60 ACLs before starting the experiment; the 1958 and 1959 data were from *S*s who had completed an average of only 11 ACLs before the dramamine assessment. Yet the results of all experiments were similar. (We have been able to elicit a steady interest and motivation from selected patients who completed ACLs daily for periods as long as two years.)

d. The ACL gives a wide range of information by including a number of affective dimensions. The range and kind of information can be altered at will.

e. The ACL gives an "instantaneous" evaluation (3 to 5 minutes for most *S*s) at a given point in time of drug action.

E. SUMMARY

This paper is the second in a series describing four drug studies. The first paper described methodology in detail. Descriptive pictures of changes in moods, emotions and motivations after drug and placebo were obtained by use of 130 word Adjective Check Lists.

1. After 100 mg of dramamine (dimenhydrinate) the *S*s were more drowsy,

sluggish, languid, withdrawn, dizzy, silent, depressed and antagonistic, and were less energetic, business-like, effective, talkative, decisive, confident, good-natured, witty and carefree. Results after 80 mg were similar but quantitatively less.

2. After 50 mg of dramamine the effects were similar except that *S*s were not withdrawn or dizzy, nor had they become less effective, talkative, confident, good-natured, witty or carefree. They were also more calm, contented and gentle, and less tense.

3. Comparison of two hour and four hour dramamine effects show a reduction at four hours and a loss of certain effects, especially physical symptoms.

4. Placebo used in the same experiments had no effects or slight effects.

5. The reliability of the adjective change scores after dramamine was measured by $^r = + .92$ between repeat administrations of the drug. The random character of the changes after placebo was confirmed by $^r = —.07$ to $+.05$ between repeat administrations of placebo.

REFERENCES

1. ALEXANDER, S. J., COTZIN, M., & WENDT, G. R. Chemical studies of behavior: I. The effects of a small dose of hyoscine on performance. *J. Psychol.,* 1955, **39**, 389-402.
2. BEECHER, H. K. The powerful placebo. *J. Amer. Med. Assoc.,* 1955, **159**, 1602-1606.
3. CHIN, H. I., NOELL, W. K., & SMITH, P. K. Prophylaxis of motion sickness: evaluation of some drugs in seasickness. *Arch. Int. Med.,* 1950, **86**, 810.
4. GAY, L. N. A subsequent report on dramamine: the prophylactic and therapeutic control of motion sickness and vestibular disturbances. *Mil. Surgeon,* 1951, **108**, 324.
5. HARWAY, V. T., LANZETTA, J. T., NOWLIS, H. H., NOWLIS, V., RIESEN, A. H., & WENDT, G. R. Chemical influences on behavior: II. Development of methods and preliminary results on the effects of some drugs on emotional and social behavior. Washington, D.C.: Office of Naval Research, U.S. Navy, Contract N6ori-126, T. O. I, 1953, 31p.
6. NOWLIS, H. H., NOWLIS, V., RIESEN, A. H., & WENDT, G. R. Chemical influences on behavior: III. The effects of dramamine and scopolamine on social and emotional behavior with comparison data on other drugs. Washington, D.C.: Office of Naval Research, U.S. Navy, Contract N6ori-126, T. O. I, 1953, 19p.
7. PAYNE, R. B. The effects of drugs on psychological efficiency. *J. Aviat. Med.,* 1953, **24**, 523.
8. ROUECHÉ, B. Placebo. *The New Yorker,* Oct. 15, 1960, 85-103.
9. WENDT, G. R. Methods of recording action. *Arch. of Psychol.,* 1938, **32**, 83p.
10. ———. III. A study of the subjective effects of small doses of benzedrine sulphate on individuals susceptible and those non-susceptible to motion sickness, including observations on psychogenic symptoms *In* Studies in Motion Sickness. Series A. Washington, D.C.: Civil Aeronautics Administration, Division of Research, Report No. 40, 1944, 45p.

11. ———. Effects of certain drugs used in self-medication in relation to driving performance and traffic hazards, *In* Health, Medical and Drug Factors in Highway Safety. Washington, D.C.: National Research Council, 1954, Publ. No. 328, 226-233.

12. ———. Methods for the Evaluation of Pharmacotherapy, Bethesda: Psychopharmacology Service Center, National Institutes of Health, 1956, 6p.

13. WENDT, G. R., & CAMERON, J. S. Chemical studies of behavior: V. Procedures in drug experimentation with college students. *J. Psychol.*, 1961, **51**, 173-211.

Department of Psychology
University of Rochester
Rochester 20, New York

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 281-287.

AN EXPLORATION OF SOME PERSONALITY CORRELATES TO BUSINESS RISK-TAKING BEHAVIOR*

Department of Psychology, Los Angeles State College

JOSEPH G. PHELAN

A. INTRODUCTION

The objectives of this research investigation consist of exploring certain personality variables as possible correlates to decision making in business risk-taking situations. The dimension with which the writer has chosen to deal is that of risk-taking behavior and its relationship to business economic decisions. The specific hypothesis states that there are significant personality differences between *S*s high and *S*s low in risk-taking behavior when measured under simulated business risk-taking circumstances. The criterion employed to assess the proposed hypothesis is the business game manufactured by Avalon-Hill Co. and called "Management." This game presents business situations wherein variables related to typical business transactions may operate and which require *S*s to make various decisions representative of those demanded of certain levels of executive management. Conditions of risk taking are defined by the operations of the game and will be elaborated upon at a later point.

It is a repeatedly indicated observation of those psychologists operating in personnel selection and analysis that individuals seeking executive management positions differ with respect to personality factors, some of which serve to enhance their effectiveness and others which serve to militate against proficient job performance. Research in this area has, until quite recently, been notably lacking. Considering its pragmatic necessity, however, it certainly appears to merit further investigation (Haire, 1959). Edwards (1954) summarized the work on decision making and its relation to economic behavior. Atkinson (1957) has considered the question of motivational determinants of risk-taking behavior while Scodel, Ratoosh, and Minas (1959) have investigated the personality correlates of risk-taking behavior. These questions are directly related to the typical business situation and thus invite an examination of psychological principles relevant to this matter.

The dimension chosen for investigation is that of economic risk-taking

---

* Received in the Editorial Office on November 7, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

decision behavior in business situations. This dimension appears to the writer as being significant because of the obvious financial risks attendant upon executive level economic decisions.

The "Cautiousness" subscale on the Gordon Personal Inventory (GPI) was employed as an independent criterion against which risk-taking behavior could be compared. Gordon (1956) indicates that it validly predicts whether individuals typically respond in a cautious or impulsive way in situations involving risk. The superior advantage of the forced-choice tetrad type of item of the GPI as opposed to the questionnaire (self-report) method is indicated by Gordon (1951) and that it is less subject to faking is confirmed by Gordon and Stapleton (1956) and Rusmore (1956).

Specifically, with respect to the present study, the "Cautiousness" subscale appears appropriate in that it purportedly measures cautiousness in terms of taking risks, especially with money, and making quick and hurried decisions as opposed to careful contemplation of all the variables. It is predicted that cautious *S*s should price products or service so low, in order to capture their share of the market, that they will lose money in the transaction. Impulsive *S*s, on the other hand, become so eager to realize a profit that they will price their product or service so high as to price themselves out of the market.

The criterion measure employed to assess risk-taking behavior as mentioned earlier is the business game. The particular advantages and content validity of business games are indicated by Andlinger (1958), Green and Sisson (1959), and Kepner and Tregoe (1960). One of the operational definitions of success in the game is a measure of the net profits. Output, as a measure of job proficiency, is the standard most frequently employed (Ghiselli and Brown, 1955).

The business game used in the present study is entitled "Management." It may be described as a simulated management training device wherein variables related to typical business situations operate. The basic paradigm of the game is composed of six phases. These are: (*a*) buying raw material units; (*b*) processing these units into finished inventory units; (*c*) selling these finished inventory units; (*d*) making loans; (*e*) purchasing capital improvements; and (*f*) property assessment. Risk-taking behavior in the business game situation was considered to occur in the stage of the game in which opposing teams were required to decide at what unit price to offer manufactured goods for sale and how many units to put on the market. The amount of this bid was comprised of the product of the monetary value per unit and the number of units available. These bids, as well as the time taken to decide upon them were recorded as criterion measures to which appropriate non-parametric tests were applied to ascertain personality group differences.

B. PROCEDURE

The entire population of 219 *S*s taking courses in the business department of Los Angeles State College during the first summer session, 1960, was administered the "Cautiousness" subscale of the GPI in order to yield a frequency distribution along this continuum. Of these *S*s who were subsequently able to participate in the experimental situation, other *S*s being unable to attend primarily because of employment obligations, eight experimental and four control groups were arranged. A *t*-test was applied to the two groups; (those able to participate in the experiment and those unable), the results indicating no significant differences at the .05 level. Thus, the 36 *S*s in the study were chosen from an eligible group of 97.

The requirements of the experimental design called for the utilization of 12 groups composed of three randomly assigned *S*s per group. These assignments were made in terms of the particular subset of 12 *S*s and not from the sample of 36. Critical cut off points were designated at and above the 86th centile for the four cautious (C) groups; at and below the 28th centile for the four impulsive (I) groups; and between and including the 41st and 59th centiles for the four neutral (N) or control groups. The particular centile scores were adopted from the normative group of college *S*s as given in the manual of the GPI (Gordon, 1956).

A standardized orientation procedure included the teaching and practice of the business game, "Management." The particular teams involved in a given experimental setting were situated each in a separate, soundproof room with adjoining door. Tape recorders were placed in both rooms in full view of the *S*s. Situation cards, which defined the market potentialities for any given month, served as initiators for the processes of the game. The aforementioned six phases were then negotiated by the *S*s, their decisions predicated upon the delimitations of these situation cards. The unit of time for the game was in months and a given month was considered to have been concluded when the six phases of the game had been completed.

A random selection of ten cards was chosen to cover the total range of months that any two teams might encounter. Chance factors accounted for a duplication of the situation cards in months one and three. Nevertheless, all teams were operating under the same cards for months 1-N in all experimental settings, N months being differentially obtained depending on the particular session involved. Response latencies, the time required for a group to report its monthly bid, were recorded for all decisions made by a group. Measurements were made of discrepancies in "finished inventory units" from the printed requirements on each situation card. These discrepancies were measured with respect to amount of money times the number of units available.

C. RESULTS

The initial *t*-test between the two groups of those who could and could not participate in the experiment yielded a confirmation of the null hypothesis at .05 level, thereby dismissing the possibility that non-participants as a group were significantly different in their scores on the cautiousness variable than participants. Data, however, indicated that the critical cut off points could have been more severely determined had the experimenter been able to draw from all 219 original *S*s. As it occurred the 97 eligible *S*s, when reduced into three subsets of 12 *S*s each for the C, N, and I groups, were not as differentiated from each other as would have been desired.

The repercussions from making a bid, one of the dependent variables, were as follows: if a team decided on too low or too conservative a bid, it captured the market but made little or no profit, in fact, often took a loss. If a team decided on too high a bid, it ran the risk of losing its chance to sell anything in that particular month. In a direct comparison of the two bids per game, (Table 1) one made by each team, C teams bid lower than N or I teams in seven months and were the same in one month (C vs. N). Chi-square was less than .02. I teams bid higher than N or C teams in eight out of eight months

TABLE 1
THE AMOUNT OF MONEY TIMES UNITS BID FOR THE SALE OF FINISHED INVENTORY UNITS AS MADE BY "CAUTIOUS," "NEUTRAL," & "IMPULSIVE" PERSONALITY TEAMS

Team		Months		
	1	2	3	4
C	5100	17000	3900	4000
I	8300	29700	6000	6600
C	6100	8700	7000	11500
N	7800	19500	7000	21600
N	5000	16000	4800	8000
I	8000	17500	6000	14000

Note: All teams are composed of different subjects.

of play. Chi-square was less than .01. In order to compare data for all groups, a common denominator of four months, the smallest amount of months played by any two teams, was utilized.

It was also hypothesized that not only would the personality variable, "cautiousness," affect the degree of risk taking, but that the time taken by groups to evaluate all variables and conditions relevant to selling decisions, would also be significantly different. The sign test yielded significance at the .05 level for the C and I teams.

In order to interpret the magnitude differences in Table I, the Wilcoxen Signed-Rank Test was performed on the data, taking the differences of the two bids per cell. The results were significant at the .02 level.

D. DISCUSSION

Personality factors have not been adequately explored as independent variables in business risk-taking situations. The real advantage of a management business game is realized if the results can adequately be incorporated into an effective training program (Green and Sisson, 1959). If in fact personality traits do play a major role in determining the effectiveness of decisions (Scodel, Ratoosh, and Minas, 1959) and (Gaudet and Carli, 1957), then analyses of the mistakes as well as the judicious decisions must be interpreted to the participants. If personality factors act as deterrents to profitable business adventures, then it seems necessary for individuals to be aware of their limitations and to subsequently modify them for more effective use. In light of Blum's (1956) complaint that tests measuring personality characteristics or adjustment for industrial selection purposes are conspicuously absent from the test batteries now in use, it is time to begin validating some of the tests that are prevalent and making necessary revisions accordingly.

Joan Guilford (1952) has demonstrated that there are significant differences in personality types between executives and supervisors. In terms of this, generalization to other employee levels would be tenuous and validation studies for each of the particular levels of interest appear to be indicated. Stern, Stein, and Bloom (1956) point out that after the actual relationships between test scores and criterion have been determined, the assessment team may erect a model and then consider the development of additional tests in order to increase the precision of the assessment, which to be sure, is one of the ultimate aims for the future.

The rationale for measuring decision latencies in this study is directly related to the typical business situation where decisions are many times expected to be made in unreasonably short time periods. The fact that *S*s had as much time as desired to make such decisions served as an ostensible advantage to the C groups. One wonders whether they would have been less effective if rushed. Impulsive participants hurried decisions irrespective of time limits; and N *S*s were somewhere intermediate to the C and I groups.

The reasons and degree to which *S*s are resistant to alter decision-making tendencies appears to be a crucial area for further investigation. In terms of the implications of this study, it appears that further examination into different personality variables which affect business risk-taking decisions might well be undertaken.

E. SUMMARY

Decision making in a simulated business risk-taking situation was explored. Thirty-six *S*s from business classes at Los Angeles State College, Summer Session, 1960, were assigned to teams of three *S*s each, to represent degrees of "cautiousness" as defined by the Gordon Personal Inventory.

Significant differences were obtained between "cautious" and "impulsive" individuals regarding the amount of money they would risk "to sell inventory on the market," and the time taken to arrive at these decisions.

The data from this study suggest the serious investigation of personality characteristics as possible correlates to business risk-taking behavior.

REFERENCES

1. ANDLINGER, G. R. Business games—Play one! *Harvard Bus. Rev.*, 1958, **36**, 115-125.
2. ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359-372.
3. BLUM, M. L. Industrial Psychology and Its Social Foundations. New York: Harper, 1956.
4. EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380-418.
5. GAUDET, F. J., & CARLI, A. R. Why executives fail. *Person. Psychol.*, 1957, **10**, 7-21.
6. GHISELLI, E. E., & BROWN, C. W. Personnel and Industrial Psychology. New York: McGraw-Hill, 1955.
7. GORDON, L. V. Validities of the forced-choice and questionnaire methods of personality measurement. *J. Appl. Psychol.*, 1951, **35**, 407-412.
8. ———. The Gordon Personal Inventory Manual. New York: World Book, 1956.
9. GORDON, L. V., & STAPLETON, E. S. Fakability of a forced-choice personality test under realistic high school employment conditions. *J. Appl. Psychol.*, 1956, **40**, 258-262.
10. GREEN, J. R., & SISSON, R. L. Dynamic Management Decision Games. New York: Wiley, 1959.

11. Guilford, J. Temperament traits of executives and supervisors measured by the Guilford Personality Inventories. *J. Appl. Psychol.,* 1952, **36**, 228-233.

12. Haire, M. Psychological problems relevant to business and industry. *Institute of Industrial Relations Reprint No. 125.* Berkeley: Univ. of Calif., 1959.

13. Kepner, C. H., & Tregoe, B. B. Developing decision makers. *Harv. Bus. Rev.,* 1960, **38**, 115-124.

14. Rusmore, J. T. Fakability of the Gordon Personal Profile. *J. Appl. Psychol.,* 1956, **40**, 175-177.

15. Scodel, A., Ratoosh, P., & Minas, J. S. Some personality correlates of decision making under conditions of risk. *Behav. Sci.,* 1959, **4**, 19-28.

16. Stern, G. G., Stein, M. I., & Bloom, B. S. Methods in Personality Assessment. Glencoe, Ill.: The Free Press, 1956.

Department of Psychology
Los Angeles State College
Los Angeles 32, California

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 289-294.

A STUDY OF TEST-RETEST RESULTS OBTAINED ON AIRCREW PROFICIENCY EXAMINATIONS*[1]

Strategic Air Command, California

KENNETH S. TEEL AND ROBERT L. PETERSEN[2]

A. THE PROBLEM

At regular intervals, the Strategic Air Command evaluates the proficiency of its bomber crews. A part of the evaluation is accomplished through the use of multiple-choice written proficiency examinations. Since these examinations are ordinarily revised only once a year, and since there is a continuous flow of crews through the evaluation program, the authors felt that there was a need for some estimate of the extent to which the security of examination content may be compromised by the transmission of information from crew members who have taken the examinations to those who have not.

Although this study deals with test-retest relationships, its major goal is to provide some indication of the extent to which the security of test material may be compromised by administration of an examination to successive groups of testees over a relatively long period of time. Such an estimate must, in this case, be made by inference, because it is based upon the assumption that the amount in which testees are able to improve their own scores, as a result of previous exposure to the test, is a reasonable index of their ability to assist others in preparing for the exam.

This problem is of some theoretical interest for many testing situations. It is of practical importance, too, in those situations common to schools and to industrial training programs in which a limited population must be tested periodically on a relatively restricted amount of subject matter. It is in just such a situation that most military evaluation units must function.

---

* Received in the Editorial Office on November 14, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This paper is based upon work performed while the authors were employed by the Air Force Personnel and Training Research Center. The opinions and interpretations contained herein are those of the authors and do not necessarily reflect the views of the Air Force.

[2] The authors wish to express their appreciation to Dr. Philip H. DuBois of Washington University and to fellow staff members for their critical review of the first draft of this paper.

B. The Sample

The sample consisted of aircrew members who were evaluated twice during the period in which the same forms of the examinations were in use. In some cases, re-evaluation was necessitated by the fact that one or more crew members failed to qualify on the first evaluation. In other instances, normal scheduling of crews resulted in some being evaluated twice during the twelve-month period.

In all, 135 test-retest scores were available for study. Of these, 23 were arbitrarily eliminated from the sample because they were obtained by individuals who had failed the first evaluation. Since disqualified crew members are probably given intensified training during the period allowed before re-evaluation, the authors reasoned that an increase in their scores on the retest might be more logically attributable to the effects of intervening training than to the effects of the first exposure to the test.

Since disqualification of any crew member requires that the entire crew be re-evaluated, the authors also considered the possibility that all members of a disqualified crew might be given intensive training in the interval between evaluations. They found, though, that to eliminate all members of any crew having one or more disqualified members would have resulted in excessive depletion of the sample.

With the elimination of the 23 disqualified cases, the test-retest sample was reduced to 112 cases distributed among 11 crew positions. Under these circumstances, no investigation of test-retest relationships by crew position was possible. Instead, all scores were converted to percentages and pooled to make a single sample of 112 cases. Comparison of the coefficients of variation computed for each of the tests indicated that the tests were sufficiently homogeneous in variance to permit such a pooling.

C. The Examinations

The examinations were of the multiple-choice, five-alternative type. They varied in length from 85 to 181 items, averaging approximately 130 items each. In all cases, they were designed to test both theoretical and practical knowledge in a variety of job areas.

Kuder-Richardson estimates of reliability for previous forms of these examinations ranged from .68 to .83. No good estimate of validity can be given, however, because of the lack of any completely satisfactory external criterion.

D. RESULTS

1. *Test and Retest Mean Scores*

The mean score on the first administration of the examinations was 65.73 per cent for the 112 members of the sample. The retest mean was 70.55 per cent—an increase of 4.82 per cent over the original testing. Since this difference is statistically significant at the 1 per cent level of confidence, the presence of some variable tending to make retest scores generally higher is indicated.

2. *Test and Retest Score Variability*

The standard deviation of the distribution obtained from the first testing was 11.45 percentage points, while that for the retest was 10.61. The difference of 0.84 falls far short of either statistical or practical significance. Therefore, it may be concluded that the data show no reliable evidence of any tendency for the retest distribution to become compressed as a result of previous exposure to the test material.

3. *Correlation between Test and Retest Scores*

A Pearson product-moment correlation of .79 was obtained between test and retest scores. Examination of the scatter plot showed a linearity of regression and a satisfactory degree of homoscedasticity throughout the range of scores. The least squares regression line for this relationship had a slope of 0.74, indicating a tendency for those with lower scores on the first testing to show more retest improvement than those with higher original achievement.

4. *First Scores Versus Improvement on the Retest*

When scores on the first testing were correlated with the amount of improvement in retest scores, a coefficient of —.42 was obtained. This coefficient, which is significantly different from zero, indicates that the magnitude of the score achieved on the original testing is a real variable with respect to the amount of test-retest difference. The slope of the least squares regression line for this relationship was —.26, indicating, as did the regression of retest scores on test scores, that improvement in retest score is inversely related to the magnitude of the original test score. Either line shows that, for scores of 50 per cent on the first test, an average increase in score on the retest of about 9 per cent would be expected. At 85 per cent on the first test, the expected retest mean would be the same; for scores above 85 per cent, a retest score lower than that made on the original test would be predicted.

The obtained negative relationship between initial score and gain is not surprising in view of the fact that the variance of the retest was essentially the same as that of the original. As a matter of fact, the only case in which a positive relationship might be obtained would be that in which the variance of the retest were substantially greater than that of the original.

5. *Elapsed Time and Test-Retest Differences*

The elapsed times between the first and second administration of the tests varied from 42 to 237 days. Analysis of these data revealed no significant relationship between the length of the time interval and the difference between test-retest scores.

E. DISCUSSION

The discussion of results is directed at two general areas: test-retest score relationships, and the gradual compromise of test material through repeated use with successive groups.

1. *Test-Retest Relationships*

The statistically significant increase in retest means may be attributed largely to learning during the test-retest interval, because all aircrew members are exposed to continuous on-the-job training. To be sure, increases in proficiency resulting from such training should be minimal when the sample is composed only of highly experienced personnel, as in this instance. Nevertheless, the fact that those who were lowest in original achievement showed the greatest improvement on the retest fits nicely with the postulation of intervening training as an effective influence on retest scores.

It has already been noted that crew members who were disqualified on the first evaluation were eliminated from the sample. Not eliminated, however, were those who were members of the same crew but who themselves were qualified on the first evaluation. Since all members of a disqualified crew are re-evaluated, and since disqualification of any crew member on the second evaluation results in reassignment of the crew as a whole, it might be expected that crew members who were minimally qualified on the first evaluation would be likely to receive more individual training than those who were highly qualified. Also, it should be noted that the low scorers have more room for improvement and are more likely to be motivated toward higher achievement on the retest, while the high scorers have little room for improvement and may perhaps suffer from some overconfidence. If so, they may engage in less preparation than for the first testing.

If one accepts this explanation as being reasonable, the compromise of test material by previous exposure would seem to be small. It is possible, of course, that some of the retest gain may have resulted from discussion of specific items immediately after the initial test. Unfortunately, there was no way of estimating the extent to which such discussions were a factor. The average number of items per test (130) and the relatively short average time limit (90 minutes), however, would seem to minimize the probability of *S*s remembering many of the items.

2. *Gradual Compromise of the Tests*

With respect to the problem of maintaining test security in a situation where the same examinations are administered to successive groups over long periods to time, the data appear to suggest that no serious amount of compromise would be expected to result from the transmission of information from one group to the next. This must remain something of a speculation, however, for the data of the present study are not directly related to this problem. Considering, however, (*a*) that high scorers on the original test did not generally improve their performance; (*b*) that the improvement shown by low scorers may be logically attributable to training rather than memory of the tests; and (*c*) that the group as a whole showed a gain of less than five per cent upon the second administration of the same tests, it does not seem likely that past testees could convey any important amount of test information to future testees. A possible exception, not covered by the data of this study, would be a situation in which a recently tested group would have the opportunity to discuss examination content with a group scheduled to be tested almost immediately following the discussions.

F. Summary and Conclusions

Examination of 112 test-retest scores obtained from aircrew proficiency examinations leads to the following conclusions:

1. A statistically significant increase of 4.82 per cent was found when the mean score for retest was compared with that for the original test.

2. Within the limits of 42 to 237 days, the amount of intervening time was found to have no significant effect on the difference between test and retest scores.

3. Level of achievement on the first test was significantly, but inversely, related to test-retest score differences. *S*s with low original scores showed the greatest improvement in retest scores.

4. The data were interpreted as suggesting that the intensive training given low scorers, along with their own increased motivation to prepare for

the retest, are more important factors in their retest improvement than is their recall of specific items from the previous testing.

5. With respect to the relationship of test-retest data to the problem of test security in a situation where the same test is administered to successive groups over long periods of time, the authors judged it unlikely that the security of test material could be seriously threatened by those unable to improve their own retest scores in any amount greater than that shown here. A possible exception to this would be where recent testees were allowed to communicate with others scheduled to take the tests in the near future.

Because of the fact that the sample was drawn from a high proficiency group and since the minimum test-retest interval was 42 days, the authors caution against extrapolation of these results to situations not directly analogous.

6511 Driscoll Street
Long Beach 15, California

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 295-299.

THE STRESSED CONDITIONING OF PSYCHOTICS*

Department of Philosophy, Tulane University

JAMES K. FEIBLEMAN

A. INTRODUCTION

In this paper an extension of existing techniques for the treatment of certain types of psychoses is proposed. Recent experiences with the victims of "brain-washing" or of "coercive persuasion" suggest a more humane medical version for the mentally ill (1, 10, 12).

In many types of psychoses it is possible to discover a delusional system. Focused on some common obsession, as when the patient thinks he is Napoleon or the victim of persecution, there will be a general rearrangement of the data of all other experiences. The most completely systematized delusions are those peculiar to paranoia, but some found in paranoid schizophrenia are sufficiently developed to be called delusional systems. A delusional system is one which is internally consistent and externally in correspondence with the contradictories of some propositions which are supported by the relevant known facts.

Elsewhere it has been argued that a delusional system may represent the substitution of a private retention schema for a public one (3). Retention schemata have been defined already as sets of beliefs maintained in the unconscious, from where they may rise to the consciousness or otherwise influence behavior (4). In the case of the psychotic, the private retention schema is a delusional system, making considerable difficulty for the patient in his efforts to adapt to his immediate environment, while his substitution of his private retention schema for the public retention schema makes him a source of considerable difficulty for society (6). Therapy in both cases is urgently indicated.

In neurotics and normal individuals the interview dialogue may be sufficient to effect the desired substitution. Traditional examples are not wanting, as those who have read the Socratic dialogues are aware, and more recent instances may be adduced (7). However, in psychotics a more extended treatment is necessary.

The proposed therapy will consist of two separate stages. First, the

---

* Received in the Editorial Office on November 27, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

delusional system will have to be dissipated, and then an appropriate substitution of the public retention schema properly effected. The first stage will be accomplished by experimental extinction, the second by stressed conditioning.

B. EXPERIMENTAL EXTINCTION OF DELUSIONAL SYSTEMS

Wolpe has shown that anxiety is a conditioned response, and moreover one which can be cancelled by aggressive assertiveness (14). In some psychotics also this may be the case. If so, experimental extinction may be effective.

In a patient with a delusional system it is possible to see the phenomenon of conditioning. Disturbed individuals were the ones most easily conditioned, and hence the ones by whom the delusional systems could be most firmly retained and recalled. The delusional system has been functioning as a public retention schema to a degree that makes the situation untenable for other individuals in the patient's immediate environment, hence the necessity for hospitalization.

The first desired effect is the dissociation of the delusional system from its security function as a public retention schema. This will have to consist in a process of experimental extinction. If every time the patient is reminded of his delusional system he is also noxiously stimulated, an experimental extinction of the delusional system can be accomplished.

Toward this end it will be necessary first by means of a series of interviews to isolate and abstract the delusional system. This can be done little by little and the parts then reassembled. The therapist will need to know the structure of the whole, but he will have a use for the parts. Most such structures take deductive form, they consist in a few axioms from which a larger number of theorems are drawn. In the therapeutic process of experimental extinction, more attention will be paid and more time given to the axioms than to the theorems.

The first part of the therapy, then, will consist in the process of experimental extinction applied to the axioms and theorems of the delusional system by means of repeated noxious stimulation. The selection of the proper agent to act as the conditioned stimulus will have to be a matter of trial and error. The noxious stimulus could be a distasteful liquid or gas, or a painful tactile stimulation. Something less severe than electroshock would be preferable, for several reasons. In the first place, the therapy will have to be repeated a number of times, and the patient will benefit more if conscious while habit patterns are being established. In the second place, electroshock is required for the positive next phase which is to consist of stressed conditioning.

The noxious stimulation is a negative substitute for the positive fact that the delusional system is losing its need-reduction value as a security system. The procedure is the same followed in the inhibitory conditioned reflex. The propositions of the delusional system will have to be repeated to the patient by the therapist, to the accompaniment of the application of whatever noxious stimulus has been chosen for the purpose. This should have the effect on the patient of reducing his dependence upon the delusional system. A gradual diminution of belief in the system should be induced. After a sufficient number of such treatments, an occasional repetition of the noxious agent functioning as a conditioned stimulus alone should be adequate for reinforcement.

The second part of the therapy is as important as the first. Without stressed conditioning, there would have been little point in practicing experimental extinction. The patient deprived of his delusional system without further interference would be left to drift back into it. At the present time, in the treatment of psychotics by means of electroshock, the first stage of a two-stage process is employed but not the second, the channels are cleared without the necessary subsequent programming. It is not advisable to remove an undesirable program without substituting a desirable one. For there is always present the possibility of "spontaneous recovery" (9, pp. 269-70) as evidenced by the phenomenon of "disinhibition," according to which a weak stimulus will restore at least partially an experimentally extinguished reflex (9, pp. 272-3).

C. THE TECHNIQUE OF STRESSED CONDITIONING

The human individual's craving for knowledge is a cortical need, having its origin in the animal need to approach and explore, called by Pavlov "the investigatory reflex" (11, p. 12). Curiosity is natural to man, and leads to belief. There is no such condition in the living individual as the absence of knowledge. Ignorance is only another name for false knowledge, the holding of beliefs which have no legitimate claim to truth. Memory traces involve the activity of a receptacle; it is not possible permanently to alter cortical tissue after a temporary process of excitation, without recording a content. So long as there are engrams, there will be beliefs.

Thus the controlled extinction of a delusional system requires the prompt substitution of a public retention schema, which must be ready to hand as a therapeutic device. An abstract formulation of a public retention schema must first be made up from a knowledge of the basic values, since public retention schema is only another name for the basic value system of a given society (2, 8, 15). For the concise purposes of stressed conditioning, it may be possible to generalize public retention schemata to a more abstract form still, and to construct an universal public retention schema (UPRS), one which would be of service in almost any society.

The UPRS would function as an emergency security system. It would provide skeletal beliefs, a structure upon which more native propositions could be hung. It would induce an active orienting response with a degree of alertness which might indicate partial recovery.

There is, however, an important warning which must be proclaimed here. For it is at this point that the dangers of indoctrination make their appearance. A misguided therapist could do incalculable harm were he alone to be assigned the task of selecting the contents of a public retention schema. The programming of such a security system would have to carry social or institutional endorsement. It is too powerful an instrument to be left to one therapist to design.

A sample formulation of the UPRS is offered here, but with the greatest reservations, more specifically with the reminder that it is intended as nothing more than a tentative proposal. The UPRS consists in five propositions, which are as follows.

1. You are not alone. 2. We are all in the same world together. 3. If you can help him (her), he (she) will help you. 4. You can help him (her) while working alone. 5. You will think of nothing that you could not think for him (her).

The technique of employing the UPRS would be to slip it into place in bits, when possible, under stress. Each administration of electroshock would be followed by the enforced rote learning of the proposition, plus the repetition of the cumulative propositions to that date. Loud sounds and other intense stimuli such as bright lights and hot air offer facilitative support to the learning process in the sense of Skinner (13, pp. 16-17). Rest and recovery periods would be occasions for bits of illustrative action, such as doing small favors for nurses and attendants.

The two steps in the proposed therapy have been outlined separately, but they would have to be carried out together. The negative step of extinguishing the delusional system and the positive step of stressed conditioning would not take place on the same occasion but would be separated by not too much time. The procedure elsewhere described as transfer matching (5) would here apply. The technique of experimental extinction would depend upon the exhaustion of the central nervous system as a result of its activity in response to repeated stimuli. The corresponding technique of stressed conditioning would depend equally upon rest and recovery occurring under the new circumstances brought about by the introduction of a fresh set of engrams.

The order of procedure would be that of primary reinforcement in the stressed conditioning we have been describing. It means following the conditioned stimulus (electroshock) with the unconditioned stimulus (knowledge). Secondary reinforcement and further stages will be furnished by the need-reduction provided by the stable conditions of the environing society. Thus it is important that as soon as appears advisable the patient be released from the hospital in the custody of his family. The feeling of security as a reinforcing agent for the beliefs which appear consistent with it and the actions which follow from it are important in establishing receptor-effector connections which it is hoped might endure.

REFERENCES

1. Biderman, A. D., & Zimmer, H. (*Ed.*), The Manipulation of Human Behavior. New York: Wiley, 1961.
2. Clough, S. B. The Basic Values of Western Civilization. New York: Columbia Univ. Press, 1961.
3. Feibleman, J. K. The cultural circuit in psychology and psychiatry. *J. Nerv. & Ment. Dis.*, 1961, 132, **2**, 127-145.
4. ———. Ecological factors in human maladaptation. *Am. J. Psychiat.*, 1961, **2**, 118.
5. ———. Transfer matching: A new method in psychotherapy. *J. Psychol.*, 1961, **51**, 411-420.
6. ———. Biosocial adaptation and mental illness. *Internat. J. Soc. Psych.* (in press).
7. ———. An illustration of retention schemata. *Psychological Record* (in press).
8. ———. Toward an analysis of the basic value system. *American Anthropologist*, 1954, **56**, 421-432.
9. Hull, C. L. Principles of Behavior. New York: Appleton-Century, 1943.
10. Lifton, R. Thought Reform and the Psychology of Totalism. New York: Norton, 1961.
11. Pavlov, I. P. Conditioned Reflexes. G. V. Anrep trans. New York: Dover, 1960.
12. Schein, E., with Schneider, I., & Barker, C. H. Coercive Persuasion. New York: Norton, 1961.
13. Skinner, B. F. The Behavior of Organisms. New York: Appleton-Century-Crofts, 1938.
14. Wolpe, J. Psychotherapy by Reciprocal Inhibition. Stanford: Univ. Press, 1958.
15. Van Zandt, R. The Metaphysical Foundations of American History. The Hague: Mouton, 1959.

Department of Philosophy
Tulane University
New Orleans 12, Louisiana

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 301-309.

CONDITIONED RESPONSES: FACT AND ARTIFACT*

Payne Whitney Psychiatric Clinic, New York Hospital and
Department of Psychology, Hunter College

SAM J. KORN AND LIVINGSTON WELCH[1]

A. INTRODUCTION

The data of contemporary psychology and psychopathology lean heavily on the conditioned response and such related phenomena as generalization and extinction. Considerable attention is focused on the representational and interpretive significance of conditioned response data, with considerations of learning theory and such problems as anxiety, nervous system types, and other psychological processes. These theoretical extensions, however, must be balanced by a continued examination of the problems of conditioned response technique. Questions concerning fundamental sources of error in procedures of conditioned response studies must be vigilantly pursued.

There are many aspects of conditioned response procedure that have to be settled by the experimenter. Some are decided by rule of thumb, some by convention, and some by convenience. Too few of the decisions concerning, for example, inter-stimulus interval, inter-trial interval, or intensity of stimuli can be based on pertinent empirical data. As a result, the consequences of variation in procedure are unclear.

This investigation concerns the problem of the adequacy of conventionally employed criteria of the *S*'s preliminary adaptation or habituation to the conditioned stimulus. Ordinarily, because any conditioned stimulus selected will initially evoke some reactions in the *S*s, conditioning (acquisition) is not begun until after the presentation of "adaptation" trials. Are the criteria of adaptation adequate? A persistent doubt that plagues serious workers in the field concerns the possibility that the facts of conditioned response data are confounded with artifacts of so-called spontaneous or unconditioned responses to the conditioned stimulus. This problem appears as a somewhat subtle, but recurrent theme in the literature. For example, two decades ago, Hilgard and Marquis (3) raised this issue in their cautions about pseudo-conditioning and sensitization. A decade ago, Hilgard again found it necessary

---

* Received in the Editorial Office on November 28, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The authors wish to express their appreciation to Ester Feinman and Ruth Sondheimer for their assistance in the laboratory.

to report that "it is often difficult to distinguish between conditioned responses and other kinds of responses that occur" (2, p. 526). He points out that "it is essential that all procedures be carefully thought out and fully reported . . . for many unexpected relations are found to hold" (2, p. 527). Unfortunately, very limited data are available that enable the experimenter to reliably determine the favorable and unfavorable conditions of conditioned response procedure.

These difficulties are compounded in studies utilizing the galvanic skin response. The GSR is often viewed as a pitfall to be avoided in conditioned response research. The researcher, to begin with, must concern himself with the problems of the adequacy and fidelity of extremely complex, sensitive recording apparatus. This is, however, only an initial deterent. The GSR also requires special attention and care because the response is non-specific, virtually omnipresent. No stimulus can be considered, in fact, neutral. Stewart, et al. (5) have recently submitted the serious charge that "so far work on GSR conditioning has dealt with the adaptation and recovery of unconditioned responses rather than conditioning of responses" (5, p. 66). Their solution to these difficulties is an attempt to develop more rigorous criteria for the delineation of conditioned responses. In this manner, errors due to the intrusion of spontaneous or orienting responses may be minimized. The same goals may be achieved through the development of new techniques for adapting *S*s to the conditioned stimuli that may be employed. It is this aspect of the problem that has been set as the task of this investigation. Two distinctive procedures of protracted presentation of a "neutral" stimulus are utilized. There will be no attempt to condition the *S*s through the imposition of an unconditioned stimulus. In this manner, the direct examination of the vagaries of adaptation of the GSR may be pursued.

B. EXPERIMENT I

1. *Procedure*

The subjects were 20 female college *S*s enrolled in introductory psychology courses at Hunter College. Their ages ranged from 17 to 19 years, and they were volunteers.

GSR apparatus consisted of a custom-made pathometer and a Texas Instrument Rectiriter which gave a paper tape recording. The GSR leads were fastened to the *S*s' palms and they were seated about five feet in front of a ground glass screen. *S*s were tested individually and alone. The room in which they sat was completely darkened.

The stimulus was an amber light (7 watt frosted bulb) projected through a 1.6 inch aperture that was covered with ground glass. Duration of the stimulus was three seconds, controlled by a Lafayette Timer. The inter-trial interval ranged from 30 to 90 seconds and was determined essentially by the *S*s' response and recovery pattern. The stimulus was presented when the *S*'s GSR was fairly level. No other systematic inter-trial interval was imposed. A total of 25 presentations of the stimulus was scheduled over a test period of approximately 25 minutes.

While the *S*s were being oriented to the laboratory, they were told that we were interested in measuring some of their reactions to stimuli, and that the galvanic skin response would be measured. The placement of the electrodes was then explained, and they were instructed to sit and pay attention to what would be going on. They were told that every once in a while a light would go on and then it would go off. All they had to do was pay attention. After they were dark adapted for about five minutes, testing began.

Responses were scored by comparing the *S*s' resistance level, in ohms, immediately prior to the onset of the signal light and the peak resistance decrease within the 7.5 seconds following cessation of light. The response equalled the post-stimulation magnitude of resistance decrease divided by pre-stimulation resistance. This multiplied by 100 yielded a per cent change, or response per cent.

2. *Results and Discussion*

One of the more conventional adaptation procedures includes the presentation of the neutral stimulus (the CS), without reinforcement, for a fixed number of trials. These may be as few as two presentations as in Phillips' study (4), but more generally they are approximately ten (1). Apparently the number is determined by the experimental program and testing time considerations. From our data, it is quite clear that this is, at best, an incomplete criterion of adaptation. The average number of responses per *S* in the last 15 trials presented is 8.7 with a median of 9.5 responses. Thus, if we assumed that the later responses were due to interpolated conditioning, *more than 50 per cent of our data would represent false-positive conditioned responses.* These would not be conditioned responses, but intrusions of original, non-conditioned responses.

In addition, the magnitude of change in response level between the first and final ten responses of each *S* is not sufficiently great to warrant confidence in this index. In three of the twenty cases, there was actually a response increase between the first and last ten stimulus presentations. In thirteen additional cases, the average response decrease for each *S* was less than 3 per cent. In only one case was the mean response decrease greater than 5.2 per cent. Thus, through the course of 25 presentations of the conditioned stimulus, the *S*s' responses fail to show a consistent decrease in magnitude. It is important to note that even this change is a function, at least in part, of the *S*s' initial response level. The higher this is to begin with, the more likely will response magnitude reflect a decrease through the course of the stimulus series.

An alternative criterion of adaptation that has been used depends on some predetermined number of successive stimulus presentations that fail to elicit a response. For example, adaptation may be assumed to be complete when three successive presentations of the CS evoke a resistance decrease, or response, of less than 1 per cent of the pre-stimulus GSR. When this criterion is applied to our data, we find that only twelve of our twenty *S*s reached this criterion within the 25 trials presented. Despite the fact that this would have eliminated almost half of our sample from a conditioning study, it still does not provide a reliable index of adaptation. The trial at which *S*s reached the criterion of three successive trials with "no response" ranged from the fifth to the twenty-fifth, with a mean of 12.6 and a median of 11. However, the continued presentation of the CS after this criterion was reached revealed that the absence of response persisted in only three of the twelve *S*s. The fourth *S* could not furnish us with more information because she did not reach the criterion until the last trial of the entire series. *In the total of 149 trials that were presented to the Ss after they reached the "criterion," the Ss failed to maintain the adaptation level 40 times.* In other words, if the post-adaptation responses had followed a conditioning procedure, 27 per cent of the responses would have been false-positive responses. They would have been confounded with the non-conditioned, original responses to the CS.

These data clearly demonstrate that the *S*s' spontaneous responses to a signal light will, most frequently, persist for 25 trials and approximately 25 minutes. Any method of inducing adaptation currently employed, through a short series of preliminary habituation trials, or waiting until the *S* fails to respond to a number of successive stimulus presentations, does not eliminate the possibility of the recurrence of responses to the CS even before conditioning training has begun.

C. EXPERIMENT II

1. *Introduction*

Some early pilot work in our laboratory has suggested that adaptation may be more consistently obtained when the habituation trials include the presentation of the CS in a matrix of similar stimuli. Experiment I revealed that the repeated presentation of a single stimulus at irregular intervals does not reliably eliminate non-conditioned GSR responses. Some earlier observations at this laboratory have suggested that these responses may be extinguished when the CS is imbedded in a series of similar stimuli that are continuously presented. Thus, if an amber light is selected as the conditioned stimulus that will be reinforced eventually, adaptation trials should include the repeated presentation of this light together with a series of other colored lights. The *S*s, during the adaptation trials, would be continuously exposed to one or another of the series of colored lights. One of these would be the conditioned, or critical, stimulus. Further research on the effects of this pattern of stimulus presentation is presented in this study.

2. *Procedure*

The *S*s were 14 female college students enrolled in psychology courses at Hunter College. Their ages ranged from 17 to 21 years and they were volunteers.

The apparatus and equipment arrangement previously employed was again used, with two major changes. The adaptation stimuli now presented consisted of four colored signal lights; amber, red, blue and white. These were approximately 15 to 20 foot candles bright at the exposure aperture. These were individually presented through a common circular aperture that was identical with the one employed in Experiment I. Duration of each stimulus as presented was 7.5 ± .25 seconds, mechanically controlled. There was no inter-stimulus interval; as one colored light went off, the next light went on. The order of colors was randomized over a series of twenty presentations, and this series was repeated until a total of approximately 200 stimulus presentations was made. The over-all period of time expended for these trials was the same as the adaptation period employed in Experiment I.

The second change in the procedure of Experiment I involved the general illumination in the *S*s' test cubicle. Rather than total darkness, the small room was now lighted by a shaded 7 watt lamp located about 16 inches above the floor, on one wall.

The instructions to the *S*, electrode placement, the measurement of responses, etc., were all maintained as in Experiment I.

3. *Results and Discussion*

For comparison purposes, the amber light will be considered the "critical" stimulus, although there was no specific reinforcement of any of the signal lights. The other three lights of the series are considered buffer stimuli and responses to these are not germane to the problem under investigation.

The data of Experiment I may serve as a frame of reference at this point. Under those conditions, the average number of responses per *S* to the last fifteen unreinforced presentations of the amber light was 8.7 with a median of 9.5. In sharp contrast to this, in the serial presentation of stimuli, the average number of responses to *all 45* presentations of the critical stimulus was only 8.5 with a median of 8. The distribution of frequencies of responses, enumerated in Table 1, ranges from a low of four, to a high of thirteen. The maximum number of responses possible, is 45. This sharp reduction in the number of responses is clear even during the initial phase of the presentations. The mean number of responses to the *first* ten presentations of the amber light was 2.2 and for nine of the fourteen *S*s, the number of responses was two or less. The *last* ten presentations of the critical stimulus evoked only an average of 1.3 responses.

TABLE 1
DISTRIBUTION OF FREQUENCY OF GSR RESPONSES TO 45 PRESENTATIONS OF THE CS

Total number per subject	Frequency
13	1
12	2
11	1
10	0
9	2
8	2
7	4
6	0
5	1
4	1
	N = 14

This low rate of responses gains further support when the magnitude of response, presented in Table 2, is examined. More than half of the total number of responses for the entire series of stimulus presentations falls at or below the 1 per cent level. This extremely low response level is fairly evenly maintained throughout the adaptation trials. There is a reduction in the number of responses between the first and last ten stimuli presented, but in both groups of trials, low response level is preponderant.

What factors account for the extremely low response frequency and magnitude manifested in Experiment II? The amber signal light was the same in both experiments, yet the response patterns evoked were so dramatically different. We can discount the possibility that the shorter response interval that had to be imposed in Experiment II accounted for the decrease in the *S*s' responses to the same stimulus. The latency of the GSR is short enough to have permitted the initiation of response within the, approximately, 7.5 second duration of the critical stimulus. This factor may account for a reduced magnitude of response, but not for the extreme sparsity exhibited. In addition, the nature of the stimulus program offers some control and/or compensation. It is quite probable that any decreases in the recorded occurrence of responses to the critical stimulus due to the shortened response interval is balanced by the necessary inclusion of responses to preceding "uncritical" stimuli that may have extended beyond the onset of the critical stimulus. The overwhelming number of trials in which no response was evident, was also complemented by the absence of responses for extremely long periods of time during the course of the experiment. Thus, the absence of responses to the amber light was not the consequence of the presence of more salient stimuli—one or more of the other colored lights presented in the series. Examination of the records revealed that *the* S*s were not responding differentially to any of the lights*. The most striking feature of the GSR records was the over-all paucity of GSR fluctuations.

TABLE 2

Distribution of Magnitude of GSR Responses

Response %	Total frequency trials 1-45	1-10	36-45
0.1-0.5	14	7	0
0.6-1.0	50	12	8
1.1-2.0	26	9	3
2.1-3.0	13	2	3
3.1-4.0	4	0	1
4.1-6.0	4	0	1
6.1-8.0	6	1	2
above 8.0	2	0	0
	119	31	18

The implications of these two studies are essentially procedural, but their relevance to the superstructure of theory based on conditioned response data is unavoidable. However, the latter problem will not be the burden of this paper.

The assumption that repeated stimulus presentation, by itself can yield adaptation that is complete and reliable is an unwarranted one and is contrary to the data obtained in Experiment I. As a result of popular assumptions, or misconceptions, along these lines, it is quite apparent that experimenters often confuse the conditioned response with the reappearance of the original, non-conditioned, or orienting responses to the stimuli. Perhaps the later responses of the *S*s, as in Experiment I, reflect more than the recovery of pre-adaptation responses. These responses may also be the consequence of new thought processes, emotional reactions, and so on. Whatever their origins, these responses must be made distinguishable from conditioned responses *per se.* The objectivity of conditioned response technique is not at stake, but the validity of inferences drawn is seriously limited.

The results of Experiment II have revealed that under certain circumstances adaptation of the GSR is not only more rapidly attained, but is also more reliably maintained; the pre-conditioning responses to the stimuli are seriously reduced and the frequency of spontaneous recurrence of these responses is dramatically curtailed. The *S*s readily adapt to the serial presentation of similar, relatively neutral stimuli, and this adaptation is surprisingly complete. Such apparent stimulus monotony should lead to the diversion of attention, the arousal of independent thought processes and emotional reactions. Since none of the *S*s fell asleep, any new, albeit trivial, stimulus that impinges on them should evoke the orienting response and its GSR component. It appears, however, that a sustained level of orientation is provided that inhibits these intrusions. It would appear that this situation has some of the torpid effects associated with hypnotic techniques; a rhythmic, monotonous pattern that inhibits the flow of associations and is instrumental in inducing the trance state. Continued study of this technique may provide detailed analyses that can serve as the bases of procedural and theoretical formulations.

D. SUMMARY

In Experiment I, the reliability of procedures conventionally employed in conditioned response studies for the pre-conditioning adaptation of the GSR was investigated, and in Experiment II an alternative habituation technique is presented. In the first experiment, an amber stimulus light was presented 25 times, at 30 to 90 second intervals to 20 female college *S*s. Adaptation, the absence of responses or the reduction of response magnitude was not reliably maintained; the GSR persisted or recurred in an irregular manner. This persistence of the non-conditioned responses, underlines the strong possibility that the facts of CR data are compounded with artifacts of the reappearance of the original responses to the "neutral" stimulus. In Experiment II, an alternative technique of stimulus presentation is employed. Fourteen female college *S*s were continuously exposed to one or another of four different signal lights. Each light, including the amber light of the first experiment, was presented for about 7.5 seconds. As one light went

off, the next one went on—for a total period of 25 minutes or so. The course of GSR to the amber light in this situation was strikingly different from the earlier findings. In this situation, the frequency and magnitude of responses was not only reduced, but this adaptation was quickly attained and reliably maintained. Although the implications of these results are essentially procedural, further research may clarify the factors underlying the phenomenon.

REFERENCES

1. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. Exp. Psychol.*, 1949, **39**, 35-40.
2. HILGARD, E. R. Methods and procedures in the study of learning. *In* S. S. Stevens (*Ed.*) *Handbook of Experimental Psychology*. New York: Wiley, 1951.
3. HILGARD, E. R., & MARQUIS, D. G. Conditioning and Learning. New York: Appleton-Century, 1940.
4. PHILLIPS, L. W. Mediated verbal similarity as a determinant of the generalization of a conditioned GSR. *J. Exp. Psychol.*, 1958, **55**, 56-62.
5. STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60-67.

Payne Whitney Clinic
New York Hospital
525 East 68th Street
New York 21, New York

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 311-317.

SYMBOLING: A KIND OF BEHAVIOR*[1]

Department of Anthropology, University of Michigan

LESLIE A. WHITE

There is a fundamental difference between the mind of man and the minds, or *mindings,*[2] of all other species; this difference is one of kind, not merely one of degree. Only man has the ability to originate and bestow meanings upon things and events and to comprehend such meanings bestowed by others.

Holy water is a good example of this kind of behavior. Holy water has a meaning that ordinary water does not have. This meaning has, of course, a value, an importance, to millions of people. Upon what does this meaning, or value, depend? Certainly not upon the physical structure or chemical composition of the fluid. It depends upon the unique ability of the human organism to originate and bestow meanings upon acts and upon objects external to itself.

Fetishes provide another example. We have a museum specimen identified as a "calcite concretion." But, in the context from which the ethnologist obtained it, it was not "just a calcite concretion"; it was the "home," or the embodiment, of a spirit. And this spirit is the creation and projection of the human mind.

Acts, singly or in rituals and ceremonies, as well as objects, may acquire meanings from the human mind. "Biting your thumb" at someone (*Romeo and Juliet,* Act I, Sc. 1), or tugging the lobe of your left ear, may be made to mean anything one wants it to mean.

---

* Received in the Editorial Office on November 29, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This essay enlarges upon a point treated briefly in my article, "The Symbol: the Origin and Basis of Human Behavior" (Philosophy of Science, Vol. 7, pp. 451-463, 1940; reprinted in slightly revised form in Leslie A. White, *The Science of Culture,* (1949; paperback, 1958). It has been reprinted also in ETC., a Review of General Semantics, Vol. 1, pp. 229-237; *Language, Meaning, and Maturity,* S. I. Hayakawa, ed. (New York, 1954); *Readings in Anthropology,* E. Adamson Hoebel *et al.,* eds. (New York, 1955); *Readings in Introductory Anthropology,* Elman R. Service, ed. (Ann Arbor, Mich., 1956); *Sociological Theory,* Lewis A. Coser and Bernard Rosenberg, eds. (New York, 1957); *Exploring the Ways of Mankind,* Walter Goldschmidt, ed. (1960).

[2] See Leslie A. White, "Mind is *Minding*" (Scientific Monthly, Vol. 48, pp. 169-171, 1939; reprinted in ETC., a Review of General Semantics, Vol. 1, pp. 86-88, 1943-1944, and in Leslie A. White, *The Science of Culture.*

Articulate speech, however, provides us with the most significant manifestation of symboling. What does *doko* mean? It is a combination of sounds, of wave phenomena in the atmosphere, or a series of marks upon a piece of paper. But human beings can give it meaning, any meaning they choose. Like the *x* in algebra, *doko* can be made to mean a dozen eggs, ten minutes, "close your eyes," or "that class of adult males of *Homo sapiens* who are between 30 and 40 years of age, have one gold incisor tooth, and play the clarinet." As Humpty Dumpty observed, in his discourse on semantics with Alice, "When *I* use a word it means just what I choose it to mean—neither more nor less."

A *symbol* may be defined as a thing or event, an act or an object, upon which meaning has been bestowed by human beings: holy water, a fetish, a ritual, a word. A symbol is, therefore, a composite of (1) a meaning, and (2) a physical structure. A symbol must have a physical form otherwise it and its meaning can not enter our experience—unless we are willing to accept the claims of telepathy or clairvoyance. But there is no necessary relationship between the meaning of a symbol and its physical basis; the relationship between the two is purely arbitrary.

The meanings of symbols cannot be grasped and appreciated (comprehended) with the senses. One cannot distinguish holy water from natural water, or discover the meaning of biting one's thumb at someone, or ascertain the meaning of *doko*, by sensory means. Symboling is that kind of behavior in which imperceptible meanings are originated and bestowed, on the one hand, and comprehended on the other. Symboling is trafficking in non-sensory meanings. And, be it repeated, no animal other than man can have, or be brought to, any comprehension of holy water or fetishes—or sin or Sunday.

But there is another class of things or events whose meanings may be related in purely arbitrary fashion to their physical forms: a green triangle may mean food to a laboratory rat, or it may mean an electric shock, or something else. But this is a *sign*, not a *symbol.*

A *sign* is a thing or event that indicates something else. There are two kinds of signs: (1) those whose meanings are inherent in themselves and their contexts (steam issuing from the radiator of an automobile, geese flying south, jaundiced eyeballs), and (2) those whose meanings are not inherent in their physical structures and situations (the green triangle that means food, the yellow quarantine flag). One learns the meanings of signs by experience (observation, the conditioned reflex). The meanings of signs which are extrinsic to their physical forms become identified with their physical forms by means of the conditioned reflex. And, having become identified with their respective physical forms, these meanings can be comprehended and differentiated with the senses: the rat comprehends the meaning of the green triangle and distinguishes it from the red circle with its eyes. How, then, are some signs to be distinguished from all symbols if the relationship between meaning and physical form is purely arbitrary?

The nature of the English language tends to confuse the issue here. There is no such thing as a *sign* or a *symbol* in the same sense that there is a *cat* and a *dog*. A cat, a dog, is an animal having certain parts and properties. But there is no thing which, in and of itself, is either a sign or a symbol. A thing or an event is a sign or a symbol only when it is in a certain context, just as a vase is a commodity, a object or art, or a scientific specimen depending upon context; or as a person is a slave only in a certain context. The fundamental distinction then is not between *a* sign and *a* symbol, but between contexts, between kinds of behavior, between "symboling" and "signing" (again the shortcomings of the English language). A thing or event, act or object, is a symbol when it is significant in the context of symboling, i.e., originating and bestowing non-sensory meanings, and in comprehending such meanings. A thing or event is a sign when its meaning is either inherent in it and its context, or has become identified through experience with its physical form and, as a consequence of this, the meaning can be grasped and appreciated with the senses. The meanings of symbols can be comprehended but not perceived; in sign behavior perception and comprehension are congruent and inseparable.

It is clear from the above that a thing or event may be significant and have meaning in either symboling or signing contexts. *Doko* originates as a symbol. But after its meaning has become identified through experience with its physical form, it becomes a sign, and its meaning is grasped with the senses: "Did you say 'pig' or 'fig'?" the Cheshire Cat asked Alice; it was a matter of hearing.

There is, then, a kind of behavior that is peculiar to an animal species, *Homo sapiens*. It consists of originating and bestowing meanings upon things and events and in comprehending these meanings. This kind of behavior should have a name. We propose the term *symboling*. A human being *symbols*, just as it performs any other function of which it is capable. It breathes with its respiratory organs, sees with its eyes, smells with its nose, and it symbols with its unique neurological equipment. The fact that we know very little about the anatomy of this function[3] does not affect by one whit the validity of the assertion just made. We may not know much about the anatomy (the structural basis and operations) of sleep, but we know that people do sleep.

Why is the word *symbol* preferable to *symbolize* to designate this kind of behavior? The answer is, because *symbolize* already has established meanings which do not include, or encompass, the meaning that we wish here to express. To symbolize is "to represent, express, or identify by a symbol, or symbols; as, a nimbus enclosing a cross symbolizes Christ" (*Webster's New International Dictionary*). To symbolize is to do something *to* something *with* something: we symbolize Christ with a nimbus enclosing a cross. But we can symbol directly and immediately (not mediately): We can symbol an object just as we can see it or feel it. We can symbol water, i.e., give it a meaning, make it holy. We can symbol a calcite concretion, i.e., give it a meaning, make it a fetish. We can symbol the sun, an owl, bathing, wine, in short, anything that we can experience. We see the world chromatically and stereoscopically because we have a certain kind of optical equipment. We symbol the world because we have a certain kind of neurological equipment. And because we symbol, we human beings can never experience the external, physical world precisely as non-human beings experience it.[4] We do not see the same sun that they do—or rather, they do not see the same sun that we do, for it is we who have added something to the purely optical experience. To us the sun is never "just a bright, shining, heat-giving object in the sky." It is the Sun Father, the Giver of Life, or a vast mass of physical particles in a process of thermonuclear transformation. No dog or ape can experience *any* of *these* suns.

It might be well, at this point, to call attention to the fact that not all of the behavior of *Homo sapiens* is *human*. Human is distinguished from non-human, and it is symboling that makes this distinction. Man is a human being because he can symbol, but much of his behavior is not-symboling and is therefore not human: coughing, scratching, walking, and so on. It goes without saying, however, that any element of the behavior of man can be placed in a context of symboling, i.e., can be symboled. A belch, for example, may be, and in some cultures is, given the meaning "I have enjoyed my host's meal"; a cough can be made to mean "pull the trigger."

---

[3] Although we know very little about the neurology of symboling, I believe we can trace its development, as a form of behavior, phylogenetically from lower forms (levels) of behavior (see Leslie A. White, "Four Stages in the Evolution of Minding," in *Evolution after Darwin*, Vol. II, The Evolution of Man, Sol Tax, ed., pp. 239-253; Chicago, 1960).

[4] "Man has, as it were, discovered a new method of adapting himself to his environment. Between the receptor system and the effector system . . . we find in man a third link which we may describe as the *symbolic system*. This new acquisition transforms the whole of human life. As compared with the other animals man lives not merely in a broader reality; he lives . . . in a new *dimension* of reality . . . . No longer can man confront reality immediately; he cannot see it, as it were, face to face. Physical reality seems to recede in proportion as man's symbolic activity advances. Instead of dealing with the things themselves man is in a sense conversing with himself. He has so enveloped himself in linguistic forms, in artistic images, in mythical symbols or religious rites that he cannot see or know anything except by the interposition of this artificial medium" (Ernst Cassirer, *An Essay on Man*, pp. 24-25; New Haven, Conn., 1944).

We would do well here to make another distinction, namely, between *human behavior* and *culture*. Many anthropologists have defined culture as learned behavior, or socially transmitted behavior. But some have seen in this definition a threat to anthropology, for if culture is behavior and behavior is the proper subject matter of psychology, what is left to the cultural anthropologist?[5] Kroeber and Kluckhohn solved this dilemma by awarding behavior to psychology, reserving abstractions of behavior for the cultural anthropologist. This had unfortunate consequences, for these abstractions became first imperceptible and eventually unreal; the cultural anthropologist was left with no real subject matter.[6]

But the solution of Kroeber and Kluckhohn was as unnecessary as it was unfortunate. The products of symboling are: acts (tipping one's hat), objects (holy water, arrow heads), concepts (sin, hot), and attitudes (loathing of milk). These products of symboling are just what they are: products of symboling. As a class of phenomena they have no name in the lexicon of science. This is indeed a remarkable fact, for what class of phenomena is more important than that which distinguishes man as a unique species? I have ventured to give these phenomena a name: *symbolates* (if an isolate is a product of isolating, may not a product of symboling be a symbolate?).[7]

Anything may be considered in any context for purposes of scientific interpretation: astronomic, physiologic, sociologic, culturologic, etc. More specifically, symbolates may be considered in either a somatic or an extra-somatic context. That is, they may be treated in terms of their relationship to the human organism, or they may be dealt with in terms of their relationships to one another. Thus, smoking a cigarette may be interpreted in terms of habit formation, satisfactions, conforming to custom, and so on. Or, we may consider it in terms of its relationship to other symbolates, such as ideas and attitudes of right and wrong, the social roles of age and sex, the distinction between the custom of cigarette-smoking (which women may be permitted to indulge in) and cigar-smoking (which may be denied to women), the relationship between cigarette-smoking and advertising, merchandising of tobacco, beliefs (assertions) with regard to the carcinogenic effect of smoking, and so on.

---

[5] "Since behavior is the first-hand outright material of the science of psychology, and culture is not . . . it is natural that psychologists and psychologizing sociologists should see behavior as primary in their own field, and then extend this view farther to apply to the field of culture also," A. L. Kroeber and Clyde Kluckhohn. *Culture: a critical review of concepts and definitions*, p. 155 (Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard University, Vol. 47, No. 1, Cambridge, Mass., 1952).

[6] See Leslie A. White, "The Concept of Culture" (American Anthropologist, Vol. 61, pp. 227-251, 1959) for discussion of these points.

[7] See White, "The Concept of Culture," p. 231.

Symbolates (instances, or products, of symboling) in a somatic context may be called *human behavior*, the scientific study of which is psychology. In an extra-somatic context, symbolates become culture, the scientific study of which is culturology. Students of words (word phenomena) have long made this distinction between somatic and extrasomatic contexts, which they have called *la parole* and *la langue*, respectively. In the somatic context, word phenomena constitute *speech*—a kind of human behavior—the scientific study of which is psychology (conception, imagination, habituation, etc.). In the extrasomatic context words constitute a *language*, and are considered, not in terms of their relationship to the human organism, but in terms of their relationship to one another. Here we have the science of linguistics, concerned with such things as syntax, lexicon, grammar and phonetics.[8]

SUMMARY

There is a kind of behavior of which man, and man alone, is capable. It consists in originating and bestowing meanings, freely and arbitrarily, upon things (objects) and events (acts), and of comprehending such meanings. These meanings cannot be grasped and appreciated with the senses. Things and events upon which meanings have thus been bestowed may be called *symbolates*. *Symbols* (symboling) are to be distinguished from *signs* (signing). Human behavior (consisting of, or dependent upon, symboling) is to be distinguished from the non-human behavior of man. Symbolates viewed in a somatic context are to be regarded as human behavior, the scientific study of which is psychology. Symbolates viewed in an extrasomatic context are to be regarded as culture, the scientific study of which is culturology. Symboling is a distinct kind of behavior, and it should be called symboling, not symbolizing.

---

[8] See White, "The Concept of Culture" for fuller discussion of these points.

Department of Anthropology
University of Michigan
Ann Arbor, Michigan

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 319-327.

BRIGHTNESS OF MONOCHROMATIC LIGHT IN SCOTOPIC AND PHOTOPIC VISION*[1]

Psychological Laboratory, University of Stockholm

GÖSTA EKMAN AND TEODOR KÜNNAPAS

A. INTRODUCTION AND PROBLEMS

The relation of subjective brightness in *photopic vision* to intensity of stimulation was previously investigated for monochromatic light in several experiments in this laboratory. In our first investigation the method of magnitude estimation was used (7). It was found that the relation was well described, for all eight wave lengths of the experiment, by a power function of the form

$$R = c(S—a)^n, \tag{1}$$

when R denotes subjective brightness and S intensity of stimulation. The parameter a was probably first introduced by one of the present authors (3): it is related to the absolute threshold for photopic vision (1, 4). The scale constant c varies with wave length. The exponent n had nearly the same value for all wave lengths, approximately 1/3, which is about the same value as has been found for white light (18, 13).

Essentially the same results were obtained in a subsequent study by means of the method of magnitude production (8). The exponents were again practically constant for all six wave lengths of this experiment, although somewhat lower, on an average a little less than 0.3. For technical reasons the effective stimulus range was somewhat shorter in this study. Since some recent investigations have shown that the exponent may vary with the stimulus range (12, 2, 17, 19), these two experiments cannot be considered strictly comparable.

The present study deals with brightness of monochromatic light, as a function of intensity of stimulation, under conditions involving *scotopic* as

---

* Received in the Editorial Office on November 29, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported by the Swedish Natural Science Research Council. The present paper is based on two unpublished research reports (Ekman and Künnapas, 1960, 1961), in which the data are presented in more detail. Part of the experimental and computational work was performed by Mr. L. Dahlstedt.

well as *photopic* vision. For purposes of comparison the experiments were performed, in other respects, under conditions which were identical as far as possible.

B. EXPERIMENTS

1. *Apparatus*

The apparatus used in the experiments has been described elsewhere (7). The subjects observed two semi-circular openings (diameter 13 mm.) in a screen (distance from the eyes 50 cm.). The intensity of the light emerging from the openings could be varied by the experimenter by a device utilizing a polarizer and two polaroids, constituting analyzers.

Monochromatic light was obtained by means of interference filters. The same six wave lengths as were investigated in the third experiment on photopic brightness were chosen for this experiment: 459, 502, 554, 595, 649, and 672 mμ. The half-band width of these filters is about 12 mμ.

In the experiments with scotopic vision a set of neutral density filters, transmitting 1.6 per cent, was used to reduce the stimulation to scotopic level.

2. *Procedure*

The experiments were performed, both in the scotopic and the photopic condition, in two parts:

In *Experiment A* the same color filter transmitted light to both openings. In each trial a standard stimulus was presented together with the comparison stimulus; the standard intensity was chosen so that its brightness was approximately in the middle of the subjective range covered by the experiment (see Figure 1). The brightness of the standard was assigned the value 10, and the subjects were instructed to estimate, for one wave length at a time, the brightness of each comparison stimulus in relation to the standard. Six comparison stimuli were used so that seven points on the curve will be obtained when the standard is included. The range of stimulation was chosen—for each condition—so that the weakest stimulus for each wave length was subjectively very weak, but clearly visible.

In *Experiment B* wave length 502 mμ was chosen as the standard with stimulus intensity 0.10 in the scotopic and 2.00 in the photopic condition. The subjects were instructed to call this standard brightness 10 and to estimate the brightness of the other wave lengths, all with the same stimulus intensity, in proper relation to this standard brightness.

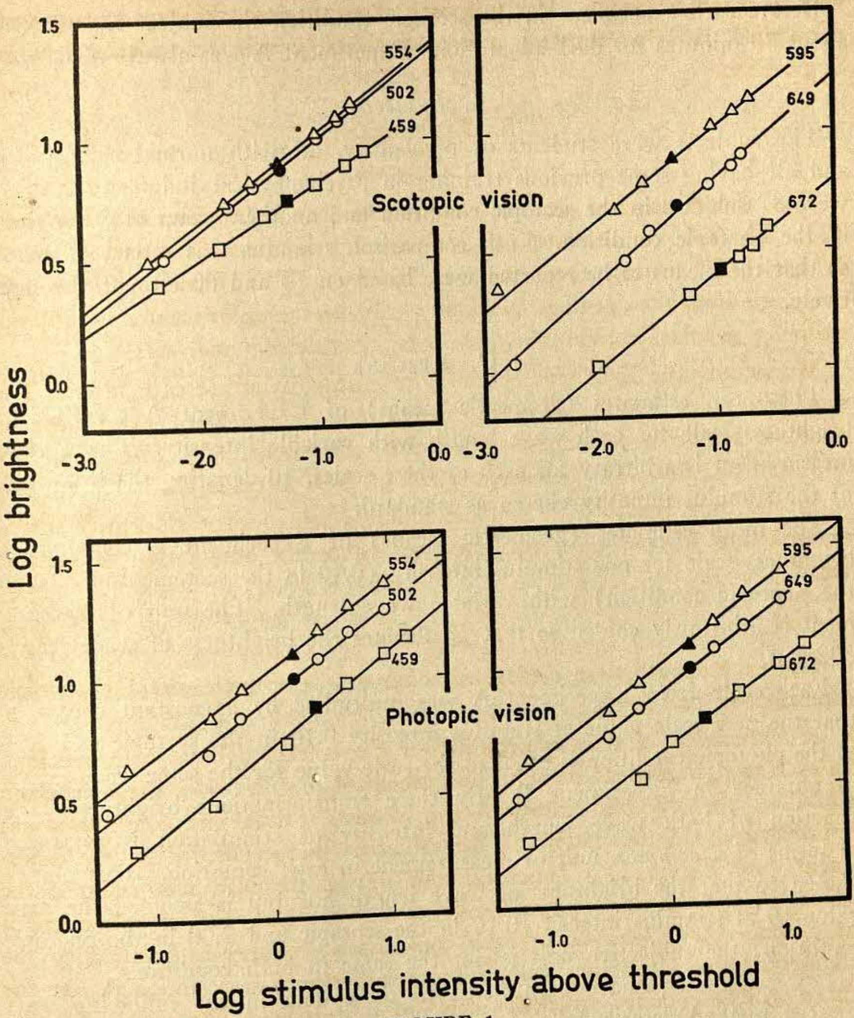


FIGURE 1

Subjective scale values for brightness plotted against stimulus values above threshold $(S—a)$. The coordinates are logarithmic so that the exponents are represented by the slope of the straight lines. The data for each condition have been divided into two separate diagrams in order to avoid confusion between nearly coinciding sets of data. The empty symbols represent experimental values and the filled symbols represent the values of the standard stimuli. All subjective scale values have been transformed to a scale with a common unit of measurement for each condition as described in the next.

Before each experiment in the scotopic condition the subject was allowed about 30 minutes for dark-adaptation. Experiment A was always made first.

3. *Subjects*

The subjects were students of psychology, all with normal color vision and all having some previous training in psychophysical judgments. There were 37 subjects in the scotopic condition and another group of 30 subjects in the photopic condition. Each comparison stimulus was estimated twice, so that the means to be reported were based on 74 and 60 estimates, respectively.

C. RESULTS

The mean estimates (geometric means) of Experiment A constitute a brightness scale for each wave length with variable intensity. The unit of measurement is arbitrary for each of these scales, 10 denoting the brightness of the stimulus intensity chosen as standard.

The mean estimates (geometric means) of Experiment B constitute a brightness scale for one stimulus intensity (0.10 in the scotopic and 2.00 in the photopic condition) with variable wave length. The unit of measurement is arbitrarily chosen so that 10 denotes the brightness of wave length 502 mμ.

Each scale from Experiment A was multiplied by a constant chosen so that the new scale value at stimulus intensity 0.10 in the scotopic and 2.00 in the photopic condition is equal to the scale value for the same wave length as obtained in Experiment B. By these transformations brightness as a function of both wave length and intensity of stimulation is measured in terms of a *common* unit of measurement in each condition. This unit is chosen so that the brightness of wave length 502 mμ is assigned the scale value 10 at stimulus intensity 0.10 in the scotopic and 2.00 in the photopic condition; the subjective unit is not the same in both conditions.

Power functions have been fitted to these data in the following way. First a method described by one of the present authors (5) was used to determine the scale constant a. Then the scale constant c and the exponent n of the power function were determined by fitting straight lines to log R and log $(S—a)$, as shown in Figure 1, where the data for each condition have been plotted in two separate diagrams in order to avoid confusion. (The exponent may be determined from the method used for determining a. This method, however, implies a treatment of linear coordinates, and since we had some special interest in the low scale values, we used the method only for determining *a* from data in the low range, and then determined the other constants from logarithmic coordinates, whereby greater weight is placed on the lower values.)

The fit of the power functions may be considered good although slight, possibly systematic deviations may be seen in the lower part of the photopic stimulus range; these deviations deserve special investigation in the transition zone between scotopic and photopic vision.

The equations are shown below:

Scotopic Vision

$R_{459} = 13.7\ (S-0.0057)^{0.31}$
$R_{502} = 24.9\ (S-0.0026)^{0.39}$
$R_{554} = 25.7\ (S-0.0039)^{0.37}$
$R_{595} = 24.9\ (S-0.0060)^{0.38}$
$R_{649} = 16.9\ (S-0.0132)^{0.42}$
$R_{672} = 6.04(S-0.0909)^{0.38}$
Average exponent: 0.38

Photopic Vision

$R_{459} = 5.17(S-0.68)^{0.39}$
$R_{502} = 8.54(S-0.46)^{0.36}$
$R_{554} = 11.17(S-0.44)^{0.36}$
$R_{595} = 10.86(S-0.44)^{0.36}$
$R_{649} = 8.36(S-0.45)^{0.36}$
$R_{672} = 4.67(S-0.94)^{0.34}$
Average exponent: 0.36

The variation of the exponent between wave lengths appears to have a random character in both conditions. The exponents for scotopic vision tend to be slightly higher (mean 0.38) than in the case of photopic vision (mean 0.36), but this difference must be considered negligible. The conclusion is that the exponent of the psychophysical power function is approximately the same for both scotopic and photopic vision. The important difference between scotopic and photopic functions is concerned with the parameter *a*, which is related to the absolute threshold and is, of course, much lower for scotopic vision. (The two conditions cannot be compared with regard to the parameter *c*, since the unit of subjective measurement is arbitrary in each condition.)

In passing it may be noted that the average exponent for photopic vision was 0.32 in the previous investigation with the method of magnitude production (8), when the power function was fitted in the same way as in the present study; both methods thus yield essentially the same results.

The main result of this investigation is the finding that the exponent is approximately the same for different wave lengths in scotopic vision, and also approximately the same as for different wave lengths in photopic vision. As a further illustration of this result Figure 2 has been constructed. It shows all the data in such a way that their trend with regard to the exponent is emphasized. For this purpose the unit of subjective measurement was chosen so that brightness is 1 when the intensity of stimulation above threshold ($S—a$) is 1. The data are plotted in logarithmic coordinates, and it is seen that they are described in good approximation, over the whole range, by a power function with the exponent 0.37, which is the slope of the straight line of the graph.

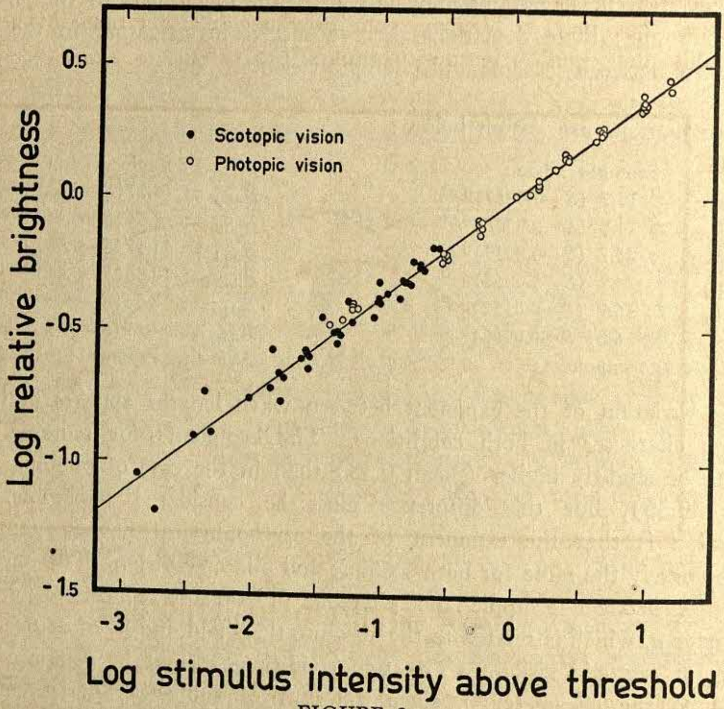


FIGURE 2

The graph summarizes all the data shown in Figure 1. The unit of subjective measurement has been chosen so that the brightness is 1, for all wave lengths, when the intensity of stimulation above threshold is 1. This means that variation of the parameter c of the power function is disregarded in order to illustrate the common trend with regard to the exponent. It is seen that all data may be described in good approximation by a power function, the exponent of which is 0.37.

The exponent of a power function may vary not only with the stimulus range but also with the stimulus chosen as standard in a magnitude estimation experiment (11, 6). Under these circumstances the numerical value of the exponent may partly be determined by the particular experimental situation but its *constancy* over wave length and between conditions may be a more general result.

The brightness data may also be represented in the form of "luminosity curves," showing brightness as a function of wave length, stimulus intensity kept constant. Measures of brightness for any stimulus intensity can be obtained from the equations above. In Figure 3 the experimental brightness values for both conditions have been plotted against wave length. The curves drawn through the two sets of points are in general agreement with inferences that may be drawn from previous determinations of absolute thresholds and equal brightness contours (15, 16).

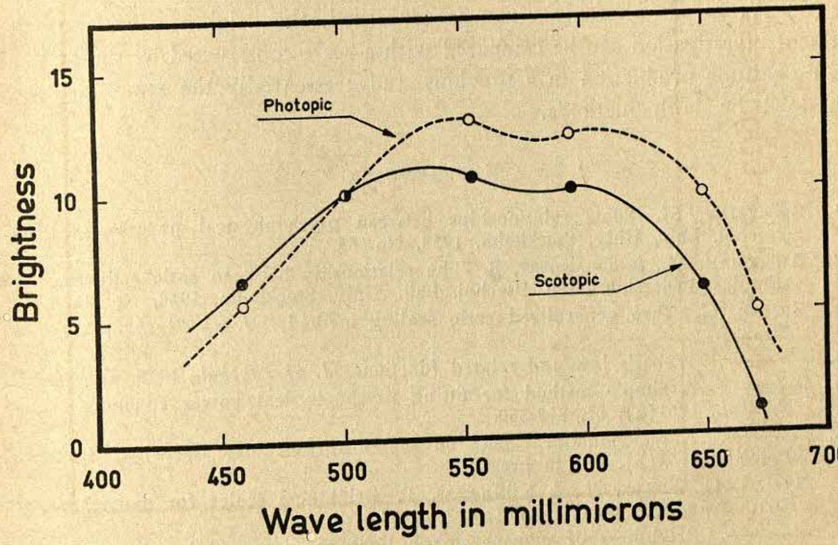


FIGURE 3
Brightness as a function of wave length. The circles represent subjective scale values, and smooth curves have been drawn to indicate the trend of these data. The units of measurement are chosen so that the brightness is 10, for each condition, at wave length 502 mμ.

It may finally be noted that inverse values of the parameter *a* of the power function (see equations) exhibit the same *general* tendency as the luminosity curves. This suggests the possibility of reducing the number of parameters necessary to describe brightness as a function of stimulus properties.

D. CONCLUSIONS

The main conclusions to be drawn from the present study when compared with certain previous studies are:

(1) Scotopic as well as photopic brightness of monochromatic light is a power function of intensity of stimulation.

(2) The exponent appears to be constant for different wave lengths under both scotopic and photopic conditions.

(3) The value of the exponent is approximately the same, probably a little above 0.35, under both scotopic and photopic conditions.

(4) The constant *a* of the power function is related both to the absolute thresholds in scotopic as well as photopic vision and to the supraliminal brightness of different wave lengths, intensity of stimulation kept constant.

(5) As far as photopic vision is concerned, the brightness scale of the present investigation can be compared with a scale constructed by the method of magnitude production in a previous study; essentially the same scale was obtained by both methods.

REFERENCES

1. BJÖRKMAN, M. Some relationships between psychophysical parameters. Rep. Psychol. Lab., Univ. Stockholm, 1958, No. 65.
2. BJÖRKMAN, M., & STRANGERT, B. The relationship between ratio estimates and stimulus dispersion. Rep. Psychol. Lab., Univ. Stockholm, 1960, No. 81.
3. EKMAN, G. Two generalized ratio scaling methods. *J. of Psychol.,* 1958, **45**, 287-295.
4. ———. Weber's law and related functions. *J. of Psychol.,* 1959, **47**, 343-352.
5. ———. A simple method for fitting psychophysical power functions. *J. of Psychol.,* 1961(a), **51**, 343-350.
6. ———. A methodological note on scales of gustatory intensity. *Scand. J. Psychol.,* 1961(b), **2**, (In press.)
7. EKMAN, G., EISLER, H., & KÜNNAPAS, T. Brightness scales for monochromatic light. *Scand. J. Psychol.,* 1960(a), **1**, 41-48.
8. ———. Brightness of monochromatic light as measured by the method of magnitude production. *Acta Psychol.,* 1960(b), **17**, 392-397.
9. EKMAN, G., & KÜNNAPAS, T. Ratio scales and category scales for brightness of monochromatic light. Rep. Psychol. Lab., Univ. Stockholm, 1960, No. 86.
10. ———. Scotopic brightness of monochromatic light. Rep. Psychol. Lab., Univ. Stockholm, 1961, No. 100.
11. ENGEN, T., & LEVY, N. The influence of standards on psychophysical judgment. *Perceptual and Motor Skills,* 1955, **5**, 193-197.
12. ———. The influence of context on constant-sum loudness judgments. *Amer. J. Psychol.,* 1958, **71**, 731-736.
13. HANES, R. M. A scale of subjective brightness. *J. Exp. Psychol.,* 1949(a), **39**, 438-452.
14. ———. The construction of subjective brightness scales from fractionation data: a validation. *J. Exp. Psychol.,* 1949(b), **39**, 719-728.
15. HECHT, S., & WILLIAMS, R. E. The visibility of monochromatic radiation and the absorption spectrum of visual purple. *J. Gen. Physiol.,* 1922, **5**, 1-33.
16. JUDD, D. B. Basic correlates of the visual stimulus. In Stevens, S. S. (*Ed.*), *Handbook of Experimental Psychology.* New York: Wiley, 811-867.
17. KÜNNAPAS, T. Scales for subjective distance. *Scand. J. Psychol.,* 1960, **1**, 187-192.

18. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. Exp. Psychol.*, 1957, **54**, 377-411.

19. STRANGERT, B. A validation study of the method of ratio estimation. Rep. Psychol. Lab., Univ. Stockholm, 1961, No. 95.

Psychological Laboratory
University of Stockholm
Stockholm VA, Sweden

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 329-347.

WHAT PSYCHOLOGISTS SHOULD KNOW ABOUT MARKETING AND ADVERTISING RESEARCH*[1]

Adult Counseling Service, Washington University

FREDERICK J. THUMIN

A. INTRODUCTION

After working in marketing and advertising research for some four and one-half years, the writer is firmly convinced that the great majority of psychologists have only the vaguest notion of what their colleagues are doing in this maverick field. A typical experimental or clinical psychologist can tell you what a personnel man does or what a human engineer does, but when either is asked what a psychologist does in marketing and advertising, the best he can manage are a few equivocal comments about repetition, subliminal perception and eye-movement cameras—all of which bear little resemblance to the actual situation.

This lack of information is a particularly undesirable state of affairs since, if the present trend continues, a substantial number of psychology graduates will enter this glamorous and lucrative field within the next few years. And, unless the academic scene changes radically, they will enter without knowing what will be expected of them, or whether they will be happy doing this kind of work. Marketing and advertising research is the youngest of all branches of psychology, and therefore very few professors have any firsthand knowledge of the subject. As things now stand, a graduate student who feels that he may wish to enter this field has a difficult time getting adequate counseling. All too often he ends up basing this important decision about his career on bits and scraps of information he is able to obtain from his professors, plus distortions and misconceptions that he gets from books like *The Hucksters* (14) and *The Hidden Persuaders* (12).

The seriousness of the problem is magnified by the fact that wise counseling is doubly important in the case of marketing and advertising research; this is a unique brand of psychology which requires a unique kind of person.

---

* Received in the Editorial Office on December 8, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The writer wishes to express his appreciation to the following individuals for their helpful comments and suggestions in the preparation of this article: M. E. Bunch, Washington University; A. G. Barclay and D. H. Kausler, St. Louis University; J. H. Leach, Gardner Advertising Company; S. R. Trottmann, Jr., Southwestern Bell Telephone Company.

Depending upon an individual's capacities, interests and personality, the work can be both stimulating and enjoyable, or it can be a miserably unhappy state of pressure and confusion.

Thus, the primary objective of this paper is to provide psychologists with basic information about a flourishing new branch of applied psychology. Hopefully, the reader will come away with a better understanding of: (*a*) what marketing and advertising research is about; (*b*) the special advantages and disadvantages it holds for psychologists; (*c*) the functions psychologists are expected to perform; (*d*) how psychologists should prepare for this field; and (*e*) the kind of person most apt to be successful and happy in this line of work.

B. Definition of Terms

Much of the confusion regarding the psychologist's role in marketing and advertising can be attributed to the fact that there are at least seven different terms which are commonly used to describe his activities; these are marketing research, advertising research, media research, copy research, creative research, consumer research and motivational research. By defining these terms at the outset, it should be possible to eliminate much vagueness and misunderstanding.

The first term, *marketing research,* is a broad one encompassing all activities designed to facilitate some facet of the marketing operation. Usually, a researcher studies one of five things—the market, the channels of distribution, the product, the package or the advertising. When advertising is the subject of an investigation, it is referred to as *advertising research.* In a general way, advertising research may be subdivided into *media research* and *copy research.* The former deals with effective allocation of the advertising budget among the various channels of communication (e.g., TV or radio? Life or Reader's Digest?), while the latter is concerned with developing and evaluating the advertising message (e.g., the best theme, layout, illustration).

Creative research is a somewhat nebulous term, used almost synonymously with copy research. Its name derives from the fact that the so-called "creative people" in an ad agency develop various advertising approaches which are then turned over to the research department for testing and evaluation.

Broadly speaking, marketing research can be conducted at one of two levels—the retail level or the consumer level. Suppose a manufacturer wants to know whether his package should be red or blue. He might attack the problem at the retail level by placing red packs in one group of retail outlets, blue packs in a matched group of retail outlets, then analyzing sales to determine which color had the more favorable influence on purchases. On the other hand, the problem might be studied at the consumer level by contacting a group of potential buyers and asking them which of the two packages they prefer. When a marketing problem is solved by going directly to the consumer and asking him questions or observing his behavior, it is referred to as *consumer research.*

The final term, *motivational research,* refers to a special kind of consumer research—that employing indirect rather than direct questioning procedures. When dealing with products which are personal in nature (e.g., underarm deodorants) or which have important psychological implications to the user (e.g., Cadillac cars), consumers are frequently either unable or unwilling to express their true feelings. In such cases, indirect methods are called for, and they may range anywhere from depth interviewing, to projective tests, to role playing (3, 8, 13).

It is perhaps unfortunate that businessmen have come to refer to this kind of investigation as motivational research, for the term often has other connotations for psychologists. Nonetheless, the term is now firmly established in marketing vernacular, and definitely here to stay.

C. Growing Opportunities in Marketing Research

The number of psychologists in marketing research has increased greatly during the past ten years, and there are strong indications that it will continue to grow at a substantial rate (2, 11). For example, just a short time ago a new division of the APA was formed—the Division of Consumer Psychology. As this paper is being written, the new division has been in formal operation less than a year, and already has attracted over 200 members. A decade ago, consumer psychologists were almost nonexistent.

Today's manufacturer, almost without exception, finds himself in a highly competitive market. Somehow he must make his product stand out from all the rest. If his advertising people fail to come up with the right combinations of words and pictures, his multi-million dollar effort may go unnoticed by a public which is currently bombarded with 12 billion dollars worth of advertising annually.

Amid such fierce competition it is little wonder that manufacturers are unwilling to "fly by the seat of their pants." Before they stake their existence on a proposed campaign, they want to be sure they are saying the right thing. An analysis of the monthly employment bulletins published by the APA and the American Marketing Association shows that a growing number of businessmen are turning to behavioral scientists to give them the assurance they need. In 1956 these bulletins carried a combined total of 23 job notices asking for psychologists (or for people trained in the behavioral sciences) to work in marketing research. During 1960, the number of such notices increased to 55—more than double the earlier figure. The following two notices from the APA *Employment Bulletin* are fairly typical of the many positions available to properly trained psychologists.

> Account Research Supervisor: New York; Ph.D. or all but dissertation for doctorate; $12,000-15,000; design, supervise and report studies in the field of advertising research including consumer and marketing studies as well as evaluation of effectiveness of advertising effort; supervise and train project supervisors and other professional research personnel; 3 to 5 years' experience in survey research desirable; consumer, marketing, copy research; public opinion sampling; strong background in Math and/or mathematical statistics highly desirable; experimental background important; operations research experience or training highly desirable; immediately. . . . (5, p. 22).
>
> Motivational Research Analyst: M.A. plus 1-2 years' experience; academic or industrial experience in consumer survey methods; minimum salary open; conducts studies on consumer attitudes and motivation, product images, packaging, etc., on cosmetic and drug products; immediately. . . . (6, p. 36).

D. How the Marketing Researcher Spends his Time

The marketing researcher performs such a wide variety of functions that it would be impossible to consider them all in a paper of this type. Thus, the discussion will be restricted to those activities which typically consume a substantial proportion of his time. The reader should bear in mind, however, that the following activities may receive greater or lesser emphasis, depending upon the particular organization with whom the individual is employed.

1. *Conducting Surveys*

A considerable amount of time is spent designing and directing surveys. These surveys are conducted for an endless variety of reasons—as endless, in fact, as the marketing problems which they attempt to solve. A few of the more common objectives are to determine: consumer attitudes toward products, advertising or companies; the socio-economic and psychological characteristics of people who purchase certain kinds of products; whether print ads (magazine and newspaper) are attracting enough readers and conveying the intended message; what a new product should be named and how it should be packaged.

As surveys differ in purpose, so do they differ in scope, methodology and the scientific rigor with which they are executed. They may, for example, be conducted locally with only 50 respondents, or they may be conducted nationally with several thousand respondents. In some surveys, respondents are interviewed by telephone, in others by mail, and in still others they are contacted personally. The sampling procedure can be as informal and subjective as having interviewers stop the first 25 people they meet on a street corner, or it can be as scientific and precise as probability methods will allow. The scope, methodology and scientific rigor of any survey is dictated by a number of factors—three of the most important being the amount of money available, the survey objectives, and the caliber of the researcher.

Psychologists who enter marketing research are sometimes surprised to learn that rarely do they conduct interviews themselves. In nearly all surveys, the actual interviewing (field work) is done by professional interviewing companies. Rightly or wrongly, data collection is regarded as a rather menial task which is not worthy of occupying the time of a well trained psychologist. There are instances, however, where the highly developed interviewing skills of the psychologist may be utilized. The most notable exception is the small group interview, often conducted as a pilot study for the purpose of developing hypotheses to be tested on a broader scale at some later date.

In most marketing surveys, the sequence of events is as follows. The researcher designs the study, develops the questionnaire, and selects the sample. He then turns the job over to an interviewing organization which collects the data in the prescribed manner. When the completed questionnaires are returned to the researcher, he assigns one or more of his clerical assistants to the task of coding and tabulating the material. When this has been done, it is the researcher's job to interpret the findings, write a formal report on the project, and make one or more verbal presentations to other members of the organization.

2. *Writing Research Proposals and Reports*

The marketing researcher spends a great deal of time writing research proposals and reports. Most sizable projects are preceded by a formal statement of the problem, method, probable cost and length of time for completion. In order to write proposals for consumer surveys, a researcher must have thorough knowledge of the various interviewing services available as well as interviewing costs. Depending upon the kind and quality of interviewing required, the average cost per interview can range anywhere from 50 cents (as in certain telephone surveys) to $15.00 or more (as in lengthy personal interviews with farmers in relatively inaccessible areas, or with busy professional men who are difficult to contact.)

When a study has been completed, a formal report is usually prepared. The format of these reports is often similar to that which psychological journals require for publication. It is especially important that the findings be reported concisely and in language that the businessman can understand.

3. *Company Meetings*

Most business executives spend an appreciable amount of time in company meetings, and the marketing researcher is no exception. It is not uncommon for him to consume an entire working day talking with other members of the organization. Meetings which require his presence evolve for a variety of reasons, a few of the most common being as follows:

a. To discuss specific marketing or advertising problems which need to be investigated. When the problem has been outlined, the researcher is often expected, spontaneously, to suggest possible means of solution. He may be asked to present a formal proposal at some later date.

b. To present research findings to appropriate members of the organization. In such meetings the researcher may be expected to explain experimental procedure, summarize results, and make specific recommendations for subsequent courses of action. He must also be prepared to defend his research, for if his findings and/or recommendations are contrary to prevailing expectations, attempts may be made to discredit the work, e.g., to question the size and representativeness of the sample.

c. To train less experienced employees. Meetings of this type take place continually on an informal basis, as when the researcher sits down with an assistant to explain certain statistical procedures which he must know in order to complete an assignment. In addition, some organizations have formal training programs which require that experienced researchers meet with younger employees on specified work days and lecture to them on relevant topics.

d. In advertising agencies, meetings are frequently called for the purpose of planning presentations to prospective clients. If, for example, an agency wishes to acquire a new cigarette account, the researcher may be asked to do something impressive in the way of research—something which will provide the agency with useful background information about the product and, at the same time, show the prospective client that the agency is both creative and industrious.

4. *Meetings With Outside Personnel*

In addition to meeting with employees of his own company, the researcher may spend considerable time meeting with outsiders. Two frequent visitors are the media representatives, who describe the unique advantages of buying space in their particular publication or time on their particular network, and the representatives of independent research firms who promote their services and facilities.

The researcher with an ad agency often has a special responsibility when it comes to meeting with outside personnel. He may be required to spend one or two days each week working directly with clients, helping them solve their marketing and advertising problems. Ordinarily, however, it is only an experienced employee who is permitted to work independently with clients, since everything he says and does reflects upon his agency. The delicate bond between manufacturer and agency may be irreparably damaged as the result of a few indiscreet comments.

5. *Analyzing Published Materials*

In order to obtain marketing information, the researcher frequently reviews and analyzes masses of published materials. For example, before a manufacturer would market a new sleep-inducing product, he would find out all he could about the incidence of insomnia, the kinds of people most apt to have sleep problems, the sales volume and advertising expenditures of similar products already on the market, and so on. Such a project may require days or weeks of reading pertinent literature in newspapers, magazines and medical journals—as well as culling facts and figures from technical and trade publications like *Sales Management* and *Drug Trade News*.

6. *Analyzing Sales*

Sales analysis and sales forecasting are nonpsychological activities which sometimes consume appreciable amounts of time. While the researcher may have clerical assistants who perform the laborious compilations and calculations, it is his responsibility to direct their activities and to interpret the final figures accurately.

7. *Traveling*

Some researchers do a considerable amount of traveling, primarily for the purpose of conducting research away from the home office. When one tries to supervise research by telephone or by mail, the risk of error (e.g., of interviewers misunderstanding instructions) is greatly increased. A researcher feels much more secure if he, or his assistant, is there in person to direct activities and to handle unforeseen problems as they arise.

To conduct research is not the only reason why a marketing researcher travels. As with any business executive, he may make trips to attend sales meetings and conventions, give speeches, meet clients—or to perform any one of a dozen other managerial functions.

8. *Specialized Research*

As a rule, researchers employed by manufacturers and by advertising agencies function very broadly in the marketing operation. There are some, however, employed primarily by independent research firms, who specialize in one particular kind of activity. A few important areas of specialization are motivational research, television ratings, visual/color research, store audits and advertising readership. By concentrating their efforts in a particular area, these researchers are often able to sharpen and refine their techniques to a considerable extent. It is these individuals who are most apt to employ special equipment, like the tachistoscope and the eye-movement camera to increase the accuracy of their measurements—as well as the salability of their service.

E. APPLICABLE PSYCHOLOGICAL SKILLS

Having discussed how marketing researchers spend their time, it is now appropriate to explicate the value of a psychological background for this type of work. Just why is it that advanced academic training in psychology is desirable? What relevant skills do psychologists have which business school graduates typically do not? The answers to these questions are provided in the following paragraphs.

1. *The Art of Asking Questions*

To a very great extent, marketing research consists of asking people questions. The answers which they give determine the manner in which millions of dollars will be spent. When their answers are misleading or incorrect, these huge expenditures are partially or totally wasted.

Since the answer to a question may be profoundly influenced by the manner in which the question is asked, marketing researchers are perpetually faced with the problem of how best to approach the crucial subject. And since any question can be asked in more ways than can feasibly be pretested, the ultimate decision is usually based on a combination of logic, experience and (especially) judgment. When a researcher is investigating the *what* question (e.g., What brand of cigarette do you smoke regularly?), there is little reason to be concerned about the precise wording of the question; the respondent is both willing and able to provide accurate information. If, on the other hand, one is interested in the *why* question (e.g., Why do you smoke Marlboro cigarettes?), subtler procedures are usually called for. When such questions are asked directly, they almost inevitably lead to comments (or rationalizations) about taste, mildness, smoothness and other semi-illusory characteristics of the product. The probability is small indeed that a respondent will admit smoking Marlboros because he satisfies certain ego-needs by identifying with the handsome, rugged, tattooed Marlboro man. Because psychologists as a group know more than anyone else about asking questions, about circumventing rationalizations and other defenses to reveal underlying motivations, businessmen have turned to them for assistance (4, 7, 10).

There is, to be sure, no single course in the graduate curriculum which teaches a psychologist how to ask questions; the skill develops gradually as the student achieves greater understanding of human behavior. When a psychologist develops a questionnaire for a consumer survey, he utilizes his entire psychological background, but especially his knowledge of interviewing, psychodynamics and projective methods. Though it is not at all necessary that the marketing researcher be a clinical psychologist, he should have had the basic clinical courses, for they increase his sensitivity to human feelings and enable him to anticipate how a person is apt to respond to a particular phrasing of a question.

2. *Experimental Design and Statistics*

One of the primary advantages which psychologists enjoy over most business school graduates is a superior knowledge of experimental design and statistics. A marketing researcher can do a more adequate job if he has a thorough understanding of reliability and validity, test and control groups, sampling error, tests of significance, correlational methods, and so on. While both qualitative and quantitative research are valuable in this field, the emphasis is usually on the latter. For example, the results obtained from depth interviews with eight or ten housewives should not be regarded as conclusive; rather they should be used to establish hypotheses which will be verified in a more extensive follow-up survey.

3. *Writing Research Reports*

By the time psychology students complete their graduate training, they have usually learned to write research reports in a reasonably clear, organized and objective fashion. The foundations of this skill are laid in the experimental lab course, and then the skill is developed more highly when the student writes his thesis and/or doctoral dissertation.

Since the marketing researcher spends considerable time writing reports in a style and format similar to that used by psychologists, this facet of graduate training is highly transferable. A psychologist is able to perform this function at a satisfactory level of proficiency almost immediately upon entering the field, whereas most business school graduates find it necessary to undergo a lengthy period of training before their writing is raised to a comparable level. It is noteworthy that the ability to write *clearly* and *correctly* is a primary qualification for marketing researchers, according to a survey conducted by a subcommittee of the American Marketing Association among commercial researchers (1).

4. *Measuring Attitudes*

Academic training in attitude measurement is especially appropriate for the marketing researcher, since he must continually assess consumer attitudes toward products, advertising and companies. Paired comparisons, the *Q* sort and the semantic differential are a few well-known psychological scales which are sometimes used to solve marketing problems. Unlike business students, most psychologists are already familiar with these and other scaling methods when they leave college.

The student who plans to enter marketing research should learn all he can about measuring attitudes, for there is still much to be done in this area. Even the more refined techniques are sometimes too insensitive to provide a really accurate picture of how consumers feel. Courses in statistics, experimental lab, test construction and industrial psychology are often instructive—but the student should also take any specialized courses which are available, such as psychophysics, logic of the social sciences, public opinion and mass communication.

5. *Electronic Computers*

The knowledge of electronic computers which some psychologists acquire during their graduate training is very useful in marketing research, for it is often necessary to process large quantities of data which could not feasibly be done by hand. While few marketing researchers operate computers themselves, it is important that they know the capacity of the machines, and how to code information for punching and tabulating.

F. WHERE PSYCHOLOGISTS ARE WEAK

The psychologist has a number of skills which are of immediate value in this line of work, but he is not without deficiencies. Consideration will now be given to them.

1. *Marketing Principles*

Upon entering marketing research, most business school graduates are familiar with marketing principles and with the kinds of problems which will confront them. Psychologists, on the other hand, are not; they usually learn marketing principles the hard way—on the job. This initial handicap can be avoided if the psychology student will take the basic courses in marketing and marketing research. As a matter of fact, marketing is the most appropriate minor for the student who plans to enter this field. If he can also find time for a course or two in advertising, so much the better; but the marketing courses are really essential. Some psychology departments offer a useful course in the psychology of advertising, but at present, such courses are few and far between.

One of the best ways for a student to familiarize himself with the marketing profession is to join the American Marketing Association. Members receive the *Journal of Marketing,* employment notices, convention news, and other useful information about the business of marketing and advertising.

2. *Sampling*

Both business school graduates and psychologists are poorly trained in population sampling. This probably comes as somewhat of a surprise to psychologists, for they generally regard themselves as quite proficient in this area; after all, they have sampled everything from white rats to college sophomores. But in marketing research, the sampling is of a very different type, and a great deal more complex. To divide 100 rats into four groups of 25 each, or to select at random 50 schizophrenics from a group of 80 is a relatively easy task. Compare this with selecting a representative sample of upper socio-economic housewives in the St. Louis metropolitan area, or a representative sample of teenage television viewers in the entire U.S. Terms like probability sample, cluster sample and quota sample are familiar to experienced marketing researchers, but foreign to most psychologists.

Unfortunately, this deficiency cannot always be overcome at the college level. A course in population sampling is sometimes offered by the sociology department or the school of business, but almost never by the psychology department. If the course is available anywhere in the university, the student would do well to include it in his program.

G. Desirable Features of Marketing Research

Every job has its desirable and undesirable aspects, but the situation is apt to be especially ambivalent for the psychologist in marketing research. In the following paragraphs the writer will present both sides of the picture —first the favorable side and then the "seamy" side. The reader should bear in mind, however, that what constitutes an advantage or disadvantage in any job situation is to some extent a subjective evaluation which may differ from one person to another.

1. *Salary*

One very desirable aspect of marketing research is that it pays top-level salaries for men of high caliber; as a matter of fact, the potential income for psychologists is probably greater in this field than in any other. A really capable young man can move from an annual salary of $6,000 to $15,000 or more within five years, particularly if he is willing to change jobs once or twice and to live in a major industrial center like New York or Chicago. In 1957, when salaries were somewhat below present levels, a survey conducted by the American Marketing Association indicated that the salaries of research directors ranged from $8,000 to more than $30,000 per year, with an average around $13,000 (15).

2. *Degree Requirements*

In most clinical, academic and governmental positions the individual is severely handicapped if he doesn't obtain his Ph.D.; this, however, is not the case in marketing research. The doctoral degree helps in bargaining for a starting salary, but beyond this point, degrees are no substitute for ability. In this highly competitive field, a very competent M.A. will soon surpass a Ph.D. of lesser caliber. As more psychologists enter marketing research, the importance of the doctoral degree will undoubtedly increase; but the emphasis will always be on ability.

3. *Stimulating Work*

The marketing business can be a virtual Utopia for the creative researcher who likes to exercise his experimental ingenuity. Almost daily, new and different problems arise which have no tailor-made solutions; and the application of psychological methods to the solution of these problems has barely begun. For a psychologist with the right kind of temperament and interests, marketing research can be as challenging and stimulating an occupation as one could hope for.

4. *Obtaining Money for Research*

It is generally easy for a marketing researcher to obtain small amounts of money (e.g., up to several hundred dollars) for the purpose of conducting projects in areas of special interest to him. If, for example, he needs $300 or so to develop a new scaling technique which he feels would be useful in subsequent projects, funds can usually be obtained. The only requirement is that the research have practical value in the business situation —that it not be conducted simply as an academic exercise.

Moreover, money for such ventures can ordinarily be obtained quickly and without having to wade through myriad red tape, as is often the case in clinical, academic and governmental positions. A verbal *O.K.* can sometimes be obtained in a matter of minutes.

H. Undesirable Features of Marketing Research

1. *Time Pressures*

More often than not the marketing researcher carries a heavy work load, and operates under rather severe time pressures. At any given moment, he may have as many as 20 or 30 projects in some phase of execution, and various persons within the organization may be pushing for rapid completion of the particular projects which interest them most. It is frequently necessary to work an evening or two during the week and a half or full day over the weekend in order to meet important deadlines. Such time pressures are especially great in advertising agencies.

The marketing researcher is sometimes frustrated that he cannot find time to conduct any special research projects of his own—even though money would be readily available. It's like having ". . . Water, water everywhere, Nor any drop to drink . . . ." He is so busy handling required projects, that to create new ones would be out of the question.

2. *Thoroughness of Work*

Because of the heavy work load and severe time pressures, one can rarely be as thorough, or even as careful about his research as he would like. There is usually the feeling that a much better job could have been done if only a few more days had been available. The experimentalist who is painfully meticulous about his research, or who is unwilling to settle for a certain degree of mediocrity in his work, is apt to find this vocation intolerable. Marketing and advertising is, of necessity, a world of deadlines and time schedules. The people in this business simply cannot afford the luxury of academic perfection. They would rather a problem be answered 80 per cent correctly today than 100 per cent correctly next month. If they wait until next month for the *perfect* answer, they may have lost their marketing advantage to competitive manufacturers.

3. *Publications*

Most marketing researchers do not have a free hand in the publication of articles, since nearly all research is regarded as confidential. When a manufacturer spends several thousand dollars to learn that one advertising campaign will be more effective than another, there is little desire on the part of management to make this information available to competitors. Before an article can be released it must be cleared by the public relations department and/or by important individuals in the organization; then, as often as not, clearance is not given. Needless to say, it is extremely difficult for a researcher to make time in an already crowded schedule to write an article which his company *may* let him attempt to publish.

If the individual remains in marketing research, a paucity of publications is no particular handicap. But if one day he wishes to enter another phase of psychology (e.g., accept an academic appointment), the publication-ghost may come back to haunt him.

4. *Nonpsychological Activities*

Psychologists in marketing research must frequently perform nonpsychological activities. Analyzing sales, establishing market potentials and setting sales quotas are just a few operations which bear little or no relation to psychology, but which sometimes consume large amounts of time. It is important for psychologists to realize that if they want to command really big salaries in this field, they must learn all phases of the marketing operation thoroughly; to know only consumer research is not enough.

As the individual ascends the pay scale and assumes positions of greater responsibility (e.g., director of research, vice president in charge of marketing), he usually drifts further and further away from psychology. The psychologist who enters marketing research must accept the fact that he is first and foremost a businessman—only secondarily a behavioral scientist.

5. *Communication Problems*

Since most businessmen are rather naive about research methodology, it is sometimes difficult for the researcher to communicate with them. When, for example, the results of a small consumer survey show that 55 per cent of the respondents prefer ad A to ad B, it may be a formidable undertaking to convince the creator of ad A that his 10 per cent advantage is not statistically significant. Nor is it uncommon for a researcher to spend considerable time developing an elaborate experimental design to solve some urgent problem, only to meet resistance from higher executives because they are unable to appreciate what is being done.

When the people who hold the pursestrings become economy minded, the control group is one of their favorite targets. It can be very difficult to convince them that a control group, which sometimes doubles the cost of the study, is necessary if the data are to be interpreted properly. The researcher often ends up using a simple approach to the problem—one which solves the problem to some extent, but not as well as a more sophisticated design would have.

6. *Female Marketing Researchers*

A final disadvantage pertains to female psychologists only. In marketing research, as in nearly every facet of the business world, it is relatively difficult for women to advance. In addition to the fact that a woman's professional career is apt to end when she marries and has children, there are the prevailing ideas that female bosses create resentment, certain jobs are too difficult for even the most competent women to handle, and women do not need to earn as much money as do men.

A combination of these factors generally prevents industry from placing women in top-level positions. Whereas really capable males can earn extremely high salaries in marketing research, equally capable females are apt to reach a sticking point around $10,000 or $12,000 per year.

I. THE MORAL ISSUE

After reading Vance Packard's best seller, *The Hidden Persuaders,* some psychologists might reject marketing research on moral and ethical grounds. This book is an excellent documentary of the position taken by the moralists, who at one time or another have argued that it is wrong for psychologists to prostitute themselves for commercial ends. Psychology, they believe, is primarily a science for healing the sick, and should not be used to manipulate consumers by appealing to their unconscious desires. Allegedly, psychologists remove the rational element from purchasing behavior and cause consumers to buy things they really don't need. Implicit is the idea that consumers may, in the long run, be turned into robots without freedom of chioce.

This viewpoint is both exaggerated and naive. In the first place, psychology is not the science of healing the sick; it is the science of behavior. And there is no good reason why it should not be used for commercial purposes, unless of course, one takes the view that commercial activity is immoral in and of itself (9, pp. 226-227). Moreover, the moralists give psychologists much undeserved credit when it comes to controlling consumer behavior. In the final analysis all psychologists do is modify a product or its image in such a way that it has greater appeal to the potential buyer. To suggest that consumers may be turned into irrational machines by exposing them to some magical combination of words in a 60-second TV commercial, is indeed stretching a point.

The idea that people should not be encouraged to buy things they really don't need, deserves special consideration. Certainly advertising whets our appetites, makes us want things, and sometimes induces us to buy things we really don't need—at least not in any absolute sense. But we are prone to overlook the fact that this perpetual discrepancy between need and desire is a natural and necessary part of our great economic system. Advertising is one very important reason why America has a higher standard of living than any other nation; it makes us continually strive for bigger and better things. Regarding the moralist's attack on material welfare, the remarks of Henry (9, pp. 227-228) are particularly enlightening:

> . . . . It is argued, for example, that people are no happier with two cars than they were with one: why try to persuade them to buy a second? This may well be true, if only we knew what we meant by 'happier': one might equally say that people are no happier with electricity than they were with oil lamps, no happier with bathrooms than they were with a zinc tub in the kitchen, and—in the ultimate resort—no happier with bronze axes than with flint ones. The people who criticize material welfare, in fact, take the standard of what they are used to as the norm of 'necessity', set arbitrary limits at this point, and insist that anything beyond this is unjustified extravagance, since people don't really *need* it.
>
> . . . . One might argue that it is a wicked waste of human effort to produce linen tablecloths, since food will be quite as nourishing eaten off bare boards . . . but this misses the whole point of the concept of *satisfactions*. There is no way out of the dilemma that either people get what they want or else what somebody else considers is good for them. . . .

In brief, it is the writer's opinion that psychologists need have no misgivings about the moral and ethical aspects of marketing research. This field may have certain undesirable features—but a lack of morals and ethics is not one of them. As a matter of fact, the popular idea that an ad man is nothing but a martini-drinking huckster in a gray flannel suit, is grossly inaccurate. This is not to say that advertising people aren't good salesmen, or that they don't drink their share of martinis. But the important thing is that they do not use salesmanship and martinis as substitutes for talent and hard work. As a group they are bright, capable and industrious people with an honest devotion to their chosen profession.

J. Personality Requirements

At this point, it is appropriate to describe explicitly the *kind* of person best suited for marketing research. Personality factors are at least as important as interest and training in determining whether one will be successful and happy in this line of work.

Certainly the marketing researcher should be a person who can perform adequately in stressful situations, for he conducts his research under stringent time pressures, and he frequently works overtime to meet unrealistic deadlines. The individual who feels strongly that research cannot be hurried, or who is compulsively meticulous about his work, has two strikes against him from the start.

He should be a man with highly developed verbal skills, and one who enjoys interacting with people. His kind of research cannot be conducted in the confines of an isolated laboratory independent of people; as a matter of fact, people are his stock in trade. He should be a good salesman, for he must continually sell his research, his products, his company and himself.

A prime requisite is a "quick mind"; he needs to discern problems with a minimum of pondering, and to suggest possible solutions on-the-spot. It is also important that he be able to draw salient facts and conclusions from unwieldy masses of data in short order. Finally, this vocation requires a particularly creative and flexible researcher—one with a special penchant for devising clever new approaches to problems.

K. Summary

The objective of this paper is to inform psychologists about marketing research, a rapidly expanding branch of applied psychology. The data provided should enable psychologists to do a more adequate job of counseling students who express an interest in this field.

A few of the major points are as follows:

a. Psychologists in marketing research spend most of their time designing and directing consumer surveys, writing research proposals and reports, analyzing published materials, and participating in business meetings.

b. Psychologists are useful in marketing research because they are trained in asking questions, designing experiments, analyzing data statistically, writing research reports, and measuring attitudes.

c. Major desirable aspects of marketing research are high salary potentials, an emphasis on ability rather than degrees, stimulating and challenging work.

d. Major undesirable features are a heavy work-load, conducting research under severe time pressures, difficulty in publishing articles, having to perform many nonpsychological activities.

e. In the writer's opinion, the moral criticisms raised against psychologists in marketing research are vastly exaggerated, and need be of little concern to persons planning to enter this field.

f. The kind of person best suited for this vocation is one who performs well under pressure, has highly developed verbal skills, has a "quick mind," is a particularly creative and flexible researcher.

REFERENCES

1. BLANKERTZ, D., FERBER, R., MULVIHILL, D. F., & ROGGENBURG, H. B. The teaching of marketing research in relation to industry needs. *J. Marketing,* 1957, **21**, No. 13, 401-412.
2. BRITT, S. H. The application of social science findings to advertising. *Amer. Mgmt. Ass. Mgmt. Rep.,* 1958, No. 15, 94-100.
3. BROWN, L. O. What motivational research is and how it works; its advantages and shortcomings. *Advertising Age,* July 25, 1955, **26**, Pp. 63 ff.
4. DICHTER, E. The Strategy of Desire. New York: Doubleday, 1960.
5. EMPLOYMENT BULLETIN, American Psychological Association. March, 1960, **13**, No. 3, Pp. 22.
6. EMPLOYMENT BULLETIN, American Psychological Association. June, 1960, **13**, No. 6, Pp. 36.
7. FERBER, R., & WALES, H. G. Motivation and Market Behavior. Homewood, Ill.: Richard D. Irwin, Inc., 1958.
8. HAIRE, M. Projective techniques in marketing research. *J. of Marketing,* 1950, **14**, 649-656.
9. HENRY, H. Motivation Research. New York: Frederick Ungar Publishing Co., 1958.
10. MARTINEAU, P. Motivation in Advertising. New York: McGraw-Hill, 1957.
11. NEWMAN, J. W. Motivation Research and Marketing Management. Norwood, Mass.: The Plimpton Press, 1957.
12. PACKARD, V. The Hidden Persuaders. New York: David McKay Co., Inc., 1957.
13. VICARY, J. M. How psychiatric methods can be applied to market research. *Printer's Ink,* May 11, 1951, **235**, Pp. 39 ff.

14. Wakeman, F. The Hucksters. New York: Rinehart, 1946.
15. Wiener, R. K. Employment outlook for marketing research workers. *Occupational Outlook Quarterly,* U.S. Department of Labor, Bureau of Labor Statistics, 1961, **5**, No. 2, 10-14.

Adult Counseling Service
Washington University
Saint Louis, Missouri

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 349-382.

RELIGIOUS ATTITUDES AT A SMALL DENOMINATIONAL COLLEGE AS COMPARED WITH HARVARD AND RADCLIFFE*

Department of Psychology, Hartwick College

HARRY K. WELLS[1]

A. PURPOSE

Early in 1947, Gordon W. Allport, Professor of Psychology, Harvard University, together with his associates, James M. Gillespie and Jacqueline Young, made a survey by means of a questionnaire of the religious attitudes of a representative sample of Radcliffe and Harvard students. The findings were published in the JOURNAL OF PSYCHOLOGY, 1948, **25**, 3-33. The investigation was undertaken, Professor Allport stated, to determine whether the oft-repeated assertion "that among young intellectuals religion is a thing of the past" was in fact the case. His general conclusion was that the assertion was essentially false, but that there was apparently a tendency, at least among Protestants and Jews, away from sectarianism, orthodoxy and fundamentalism toward a more individual, ethical and ecumenical approach (1).

This conclusion, together with the more detailed results presented in a series of tables, was qualified by Professor Allport and his associates as follows:

> "On the whole we believe that our findings are not likely to differ significantly from those that would be obtained in state universities throughout the country. They would almost certainly differ from responses coming out of small denominational colleges . . ."

It is this latter statement that prompted the present investigation.

Hartwick qualifies as "A Small Denominational College." Is it a fact that the religious attitudes of Hartwick College students "almost certainly

---

* Received in the Editorial Office on December 15, 1961, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The help of the following students, members of the author's course in "The Psychology of Religion," was absolutely indispensable:

Diane Lee Bloemer
J. Wellington Fitch
David H. Smith
Donald O. Shaver
Roger Ribreau
Paul Bernstein
David G. Terpening
Kenneth S. Roberts
Susan H. Bartlett
Susan Butts
Lillian V. Mickle

differ" from those of Radcliffe and Harvard? To discover the answer to this question, the present investigation was undertaken.

There is an important qualification that should be made with regard to this study. In the comparison with the Harvard-Radcliffe findings there is a time lapse of some fourteen years. During this period there has been, according to many authorities, a considerable growth in religious affiliation. This factor must assuredly be taken into consideration in drawing conclusions from the comparisons presented in the juxtaposed tables that follow. In this connection, however, the following statement of Professor Allport is worthy of note ". . . . the dissaffection of modern youth is probably no greater today than it was fifteen or twenty years ago (1, p. 45)." If this is true for the years prior to 1947, perhaps it holds likewise for the years following that date, a contention which would tend to minimize the fourteen-year lapse between the two surveys.

B. Method

1. *The Questionnaire*

The instrument for the investigation of religious attitudes of students at Hartwick College was the questionnaire originally employed in the survey of beliefs at Harvard and Radcliffe in 1947. In the report of their findings, Professor Allport and his associates speak of questionnaires in general as a "traditional" but "far from perfect device (2, p. 4)." They go on to say that "Questionnaires differ enormously in their quality" and that "to construct a 'good' one is a difficult task." Of their own endeavor they say "with the aid of pre-testing we believe that most potential ambiguities were eliminated from our instrument (2, p. 4)." The Allport questionnaire was considered by the present investigators to be of high quality, and it was therefore accepted with only very minor revision of wording, addition of a few more explicit multiple choice answers, and deletion of two or three phrases which were felt to be prejudicial in eliciting desired responses. For example, the opening statement of answer (6) of Question number 14, reading "The universe is merely a machine" was deleted, as was also the opening statement of answer (c) of Question number 17, reading "Religion, as Karl Marx said, is the opiate of the people." In neither case was the phrase germaine to the answer, but served, in our opinion, to load the possible answer by use of "color" words and concepts.

Two major changes were made in the original questionnaire, one at the beginning and the other at the end, both being dictated by considerations totally unrelated to the excellence of the instrument. On the one hand, for intra-mural reasons more detailed information was sought in Questions numbered 1 and 2. Data on class, major, home environment, parents' education, and parents' religious affiliation were desired for purposes of correlation, significant within the college community. On the other hand, Question number 18 of the original questionnaire, concerning the attitude of veterans of World War II, no longer has any relevance and was therefore dropped.

2. *The Sample*

Our material was obtained from 77 regular Hartwick College students attending the August Summer Session, 1961. This constituted almost an exhaustion sample of the Summer School, the total enrollment of which was 91.[2] The sample constitutes over 10 per cent of the entire-regular session enrollment of the college. This compares favorably with the quantitative adequacy of the Harvard and Radcliffe samples. (See Table A).

TABLE A

COMPARATIVE QUANTITATIVE ADEQUACY OF SAMPLES

	Total registration[3]	Sample	Per cent
Hartwick	753	77	10.22
Harvard	5435	414	7.61
Radcliffe	889	86	9.67

A comparison of the composition of the samples in terms of age and marital status, and including the special situation in 1946-47 in which veterans of World War II composed 70.3 per cent of the Harvard sample, reveals some interesting phenomena. (See Table B).

Comparing the Hartwick sample with the Harvard non-vet sample and the Radcliffe sample, it is apparent that both the age-level and the proportion of married students of the Hartwick sample are considerably higher. In the 21-or-over category, 45 1/2 per cent for Hartwick men and women as opposed to 6 1/2 per cent for Harvard non-vets and 6 per cent for Radcliffe students; in the married category, 16.9 per cent for Hartwick men and women as opposed to 0.0 per cent of Harvard non-vets and 2.3 per cent of Radcliffe students. The higher age level of Hartwick students may reflect a lower economic status of the students leading to delays in college education, while the higher proportion of married students may reflect both the higher Hartwick age level and the national trend in recent years toward marriage at a younger age, including marriage during the course of college careers. The significance of these data for the study of religious attitudes stems from the Allport thesis that age and marital status affect such attitudes. The presence of older and married students in a sample, according to this contention, would tend to increase the percentage of positive religious orientations. This is at least a factor to be kept in mind with regard to the comparison of the Hartwick sample with the findings at Harvard and Radcliffe.

---

[2] The fourteen registrants not represented in the sample were not regular-term Hartwick College students.

[3] Total registration figures for Harvard and Radcliffe are for the college year 1946-47, the year in which the Allport survey was made. The Hartwick total is for 1960-61.

TABLE B
COMPARISON OF SAMPLES BY AGE AND MARITAL STATUS
(In Percentages)

	Age		Marital status		Entire sample
	20 or under	21 or over	Single	Married	
Hartwick (Male—N 53)	58.5	41.5	83.1	16.9	100.0
Hartwick (Female—N 24)	58.4	41.6	83.4	16.6	100.0
Hartwick (Total—N 77)	54.5	45.5	83.1	16.9	100.0
Harvard (Total N 414)	46.5	53.5	93.0	7.0	100.0
Harvard (Veterans N 291)	26.5	73.5	90.0	10.0	70.3
Harvard (Non-vets N 123)	93.5	6.5	100.0	0.0	100.0
Radcliffe (Total N 86)	94.0	6.0	97.7	2.3	100.0

The qualitative comparability of the Hartwick and Harvard-Radcliffe samples is, of course, difficult if not impossible to assess. There is, however, a liability on each side which may tend to offset one another. On the one hand, the Hartwick Summer School includes a relatively large proportion of "make-ups" and "speed-ups" and therefore tends to reflect in part at least two extremes of the regular student-body. On the other hand, in the Harvard and Radcliffe sample "The students were all members of an introductory course in Social Relations, entitled *Social and Psychological Foundations of Behavior* (2)." A sample composed of members of one course is perhaps as limited as a summer school enrollment reflecting in part lower and higher extremes of student achievement. At any rate, the Hartwick sample cut across the curriculum from music and business courses to courses in the sciences and religion. Finally, the special situation prevailing at Harvard, as elsewhere, in 1946-47, the two-thirds veteran composition that is, tended to increase the non-religious aspect of the findings of the Allport investigation. Thus Professor Allport reported that "veterans seem less religiously inclined than non-veterans" and that "The difference between male veterans and male non-veterans is almost as great as that between the sexes (2)." This factor must clearly be taken into consideration in any comparison between the findings in regard to the Harvard and Hartwick samples.

In our judgment the Hartwick, Harvard and Radcliffe samples are essentially comparable with regard to quality as well as relative quantity.

C. RESULTS

1. *How do Denominational College Students Compare With Harvard and Radcliffe Students in Regard to Felt Need for Religion?*

The relevant Question is number 3. Table 1 states the question and presents the results obtained at Hartwick, Harvard and Radcliffe.

TABLE 1

"Do you feel that you require some form of religious orientation or belief in order to achieve a fully mature philosophy of life?"
(In Percentages)

	Hartwick male (N 53)	Hartwick female (N 24)	Hartwick (total 77)	Harvard (N 412)	Radcliffe (N 85)
Yes	75.5	96	82	68	82
No	19	4	14	19	12
Doubtful	5.5	—	4	13	6

The finding here is first, that roughly seven out of ten Hartwick and Harvard *men* feel a need for some kind of religious orientation while two out of ten feel no such need. The latter is exact, being 19 per cent "no" in each case, while in the former the actual percentages are 68 per cent "yes" for Harvard and 75.5 per cent for Hartwick men. There is, then, little essential difference between the felt religious needs of men at a small denominational college and at a large secular university, at least as indicated by the two samples. As a matter of fact, the difference in the "yes" answers of Hartwick men and Harvard is almost fully accounted for by the lower percentage of "yes" among Harvard veterans. Second, while the results at Hartwick corroborate the Allport finding that women are more religious than men, the percentage of women who report the need for a religious orientation is considerably higher at Hartwick than at Radcliffe, 96 per cent as opposed to 82 per cent, a difference of 14 per cent. The figures for no felt need of religion are equally striking, 4 per cent for Hartwick women as against 12 per cent for Radcliffe. A similar picture is presented by the percentage difference between the religious needs of men and women at Hartwick and the percentage difference at Harvard and Radcliffe; 20.5 per cent as against 14 per cent.

While approximately seven out of ten *men* at Harvard and Hartwick feel the need for a religious orientation and two out of ten do not feel this need, almost ten out of ten *women* at Hartwick feel such a need as against eight out of ten at Radcliffe.

The question in regard to need for religious orientation does not imply the *kind* of religion involved. Only on the basis of other questions tabulated in the following tables can such data be obtained. The present question indicates only that, given a chance to define religion in any manner they wish, roughly seven out of ten men and eight to ten out of ten women, whether in denominational or large non-sectarian universities "regard themselves as actually or potentially religious," as Professor Allport stated it.

Two conclusions can be drawn from the above. The first is in full agreement with that made by Professor Allport and his associates: "Thus it is clear that the majority of young intellectuals—ordinarily regarded as the most 'emancipated' and 'religiously radical' element in the population—definitely consider the religious sentiment as appropriate to their own developing personalities (2, p. 10)." The second conclusion is half in agreement, and half not in agreement, with the Allport thesis that his findings "would almost certainly differ from responses coming out of small denominational colleges." This thesis holds markedly for women students but not significantly for men, at least on the basis of the first question, which Professor Allport calls "The pivotal question in our scale (2)."

2. *The Influence of the Student's Own Religious Background at Hartwick, Harvard and Radcliffe*

Table 2(a), based on Question 4, presents the degree of religious influence in the upbringing of Hartwick College students as compared with Harvard and Radcliffe.

The comparison in regard to degree of religious influence in upbringing between the Hartwick and the Harvard-Radcliffe samples presents a mixed picture. On the one hand, the fact that all Hartwick students reported some degree of religious influence while 6 per cent of Harvard and 10 per cent of Radcliffe reported no such influence would indicate that there was more religious influence in the upbringing of Hartwick men and women. Supporting this interpretation is the 24 per cent to 16 per cent comparison of Hartwick men with Harvard in the "very marked" category, and the 62 per cent to 31 per cent comparison of Hartwick women with Radcliffe in the "moderate" category. The contrary is found in the 25 per cent to 29 per cent comparison in the "very marked" category of Hartwick women with Radcliffe, and the 41 per cent to 43 per cent comparison in the "moderate" category of Hartwick men with Harvard. The difference in the contrary figures is considerably less than the difference among those supporting the contention that there was more religious influence in the upbringing of the Hartwick than of the Harvard-Radcliffe sample. At any rate the Hartwick findings in this connection certainly reinforce the Allport conclusion "that religious training for college youth is by no means a thing of the past as it is often supposed to be (2, p. 11)." If anything, the results of the present study indicate that this is even more true of students at the small denominational college than at the large secular university. The difference, however, is neither as large nor as striking as might perhaps have been expected.

TABLE 2(a)
"DEGREES OF RELIGIOUS INFLUENCE IN UPBRINGING"
(In Percentages)

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
Very marked	24.5	25.0	24.5	16.4	29.0
Moderate	41.5	62.0	48.0	43.7	31.5
Slight	32.0	17.5	24.5	33.8	29.0
None at all	0	0	0	6.1	10.5
Blank	2	0	1	0	0

Table 2(b) presents a comparison of the Hartwick and the Harvard-

TABLE 2(b)
"RELATIONSHIP BETWEEN NEED FOR RELIGION AND DEGREE OF RELIGIOUS INFLUENCE IN UPBRINGING"

Percentage of degrees of religious influence reporting felt need for a religious orientation.

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
Very marked	92	100	95	82	96
Moderate	83	93	87	78	85
Slight	54	100	60	52	76
None at all	0	0	0	32	44

Radcliffe samples in regard to a correlation of the degree of religious influence in upbringing with the student's felt need for a religious orientation (a correlation, that is, between Questions 4 and 3 of the questionnaire).

The table indicates a fairly close comparison between the Hartwick total and the Harvard and Radcliffe findings in each of the degree categories, and even closer like-comparison between Hartwick men and Harvard. The comparison between Hartwick women and Radcliffe are not too far apart in the "very marked" and "moderate" categories. The one startling figure is that of the Hartwick women in the "slight" category, both in its own right and in comparison with the other four samples in the same category. Here we see that 100 per cent of the Hartwick women who report a slight degree of religious influence in their upbringing at the same time report a felt need for a religious orientation of some kind. This compares with 76 per cent for Radcliffe and 54 per cent and 52 per cent respectively for Hartwick men and Harvard. Here we begin perhaps to discern a triple tendency which will bear watching throughout the remaining tables. On the one hand, Radcliffe women, thus far at least, appear to be more religious than both Hartwick and Harvard men, even markedly so. On the other hand, Hartwick women appear to be considerably more religious than Radcliffe women. Hartwick men would at this point seem to fall somewhere between Harvard and Radcliffe, perhaps closer to the former than to the latter.

With the single exception of the Hartwick women in the "slight" category, our findings support the Allport theses, one, that "students *trained* in religion find that they *need* religion more often than do others" and, two, that "the numbers reporting the need are in direct proportion to the degree of the influence (2, p. 11)."

In Question 6 of the questionnaire students were asked to check those influences which they thought might have positively influenced their religious views. Table 3 presents the Hartwick findings as compared to those of Harvard and Radcliffe.

The comparative findings in regard to influences leading toward a religious orientation reveal a number of significant differences and only two essential similarities. Taking the latter first, the Hartwick sample closely parallels the Harvard-Radcliffe results in the "reading" and "sex-turmoil" categories. Reading plays a fairly large role as a positive religious influence for both sets of samples, while sex-turmoil plays a relatively insignificant part. In these two respects, students of the small denominational and the large secular schools agree. But there the agreement ends. Most significant among the differences are the considerably greater roles played in the small denominational college by "parental influence," "conformity with tradition" and "church teachings." Scarcely less significant are the considerably smaller roles played in the denominational college by "fear or insecurity," "aesthetic appeal" and "mystical experience." In two other categories the denominational college shows a higher percentage than the secular university, "personal influence of others" and "studies." The latter may well reflect the fact that at denominational colleges, courses in religion are an important and integral if not required part of the curriculum.

TABLE 3

PERCENTAGES REPORTING VARIOUS TYPES OF INFLUENCES

"If you are, or have been, religiously orientated, which factors in the following list do you recognize to have been contributing reasons? *Check as many as apply.*"

	Hartwick male (N—52)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N—414)	Radcliffe (N—86)
Parental influence	75	92	81	51	34
Fear or insecurity	13	38	21	43	46
Personal influence of others	49	58	51	36	43
Conformity with tradition	42	46	43	35	27
Aesthetic appeal	19	33	23	25	41
Church teachings	49	54	51	24	34
Gratitude	6	46	18	23	42
Reading	23	25	23	20	24
Studies	32	67	43	17	19
Sorrow or bereavement	6	38	16	17	26
A mystical experience	6	13	8	10	21
Sex turmoil	4	8	5	8	8

In two categories Hartwick follows the pattern of Harvard and Radcliffe but in a heightened manner. While Hartwick men follow the Harvard pattern by reporting lower roles than Hartwick women and Radcliffe for "gratitude" and "sorrow" they do so in a striking fashion. In the case of "gratitude," 6 per cent as against 46 per cent for Hartwick women and 42 per cent for Radcliffe; and in the case of "sorrow or bereavement," again 6 per cent as opposed to 38 per cent for Hartwick women and 26 per cent for Radcliffe. The respective Harvard figures were 23 per cent and 17 per cent.

Perhaps the most striking figure in the table is the 92 per cent of Hartwick women listing parental influence as a contributing reason for their religious orientation. This becomes more striking in view of the fact that Radcliffe students reported less parental influence than Harvard, 31 per cent to 54 per cent. Of this anomaly Professor Allport wrote "Just why women report less parental influence than men is not clear (2, p. 13)." In the Hartwick sample we do not have that problem.

3. *How do Denominational College Students Compare With Students of the Secular University in Regard to Remaining in the Faith of their Upbringing?*

The topic here concerns the consistencies and shifts in allegiance of students in regard to childhood and college-level church affiliation or choice. The findings are obtained by a correlation of Questions 3a and 4a of the questionnaire. Question 3a asked the students to indicate which, if any, of certain great religious systems they feel satisfactorily meets their own present religious needs. Question 4a asked in which of these religious systems they were brought up. The comparative findings, among the most interesting of our investigation, are presented in Table 4.

A number of important facts and comparisons emerge from this table.

a. All five samples indicate that there is far less shift in allegiance among Roman Catholics than among Protestants. As a matter of fact among Hartwick women and Radcliffe students who are Roman Catholics there is no shift whatever. Among Hartwick men and Harvard students who had a Roman Catholic upbringing there is in each case a slight and almost identical shift in allegiance.[4] The category "Anglo-Catholicism and Eastern Orthodoxy" presents a similar picture, with the sole exception of Hartwick women. Here however the shift, 9 per cent, is *toward* rather than *away* from Anglo Catholicism, and therefore is literally the exception that proves the rule. In general, the results of the Hartwick College investigation confirm Professor Allport's pertinent conclusion that "The more orthodox Catholic faiths lose relatively fewer adherents than does Protestant Christianity or Judaism." This finding holds as firmly for the small denominational college as it does for the large secular university.

b. In the category "Protestant Christianity," Table 4 reveals not only little significant difference, but essential agreement between the Hartwick sample and the Harvard-Radcliffe sample. Both indicate a high percentage of shift. The percentage difference between Protestant background and

---

[4] Here the term "shift" in no sense implies a movement *toward* any particular faith, but simply a *change away from* the affiliation of family background, i.e., upbringing.

TABLE 4

Comparison of Religious Background and Present Choice Among Hartwick, Harvard and Radcliffe Students

	Hartwick males		Hartwick females		Hartwick total		Harvard		Radcliffe	
	Background	Choice	Background	Choice	Background	Choice	Background	Choice	Background	Choice
Roman Catholicism	21	17	17	17	19	19	16	11	14	13
Anglo-Catholicism and Eastern Orthodoxy	6	7	—	9	4	7	4	2	2	3
Protestant Christianity	58	36	79	54	64	42	44	18	40	28
Liberalized Protestantism	2	—	—	—	1	—	7	6	9	9
Ethical, but not theological Christianity	2	5	4	4	3	4	4	11	5	5
Some form of Judaism	8	6	0	0	5	4	17	6	18	8
Other	4	5	—	8	3	5	8	5	12	8
New type needed	—	8	—	4	—	—	—	11	—	11
None	—	—	—	—	—	—	—	17	—	8
Doubtful about need	—	—	—	—	—	—	—	13	—	6
Blank	3	16	—	—	3	11	—	—	—	—

present choice for the breakdown by college of the two sets of samples is shown in the following subtable:

COMPARATIVE PERCENTAGE OF SHIFT AMONG PROTESTANTS

Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
22	25	22	26	12

The shift among Protestants represents for the greater part movement from one Protestant denomination to another. To a certain extent, however, it would appear to mean also a limited shift toward forms of Christianity other than the traditional denominations, and toward new and non-theological types of religious belief.

c. The latter contention is reinforced by the significant if limited shift of Hartwick men and Harvard students toward Ethical but not theological Christianity or in some instances Judaism-Christianity, as in the Society for Ethical Culture. As a matter of fact, "Ethical but not theological Christianity is the only religious system," as Professor Allport concluded, "to register an appreciable gain (2, p. 13)." This is true of the Hartwick male sample as well as of the Harvard sample, respectively 2 per cent to 5 per cent and 4 per cent to 11 per cent. Thus the small denominational college and the large secular university reveal the same trend, although in a less pronounced manner.

d. With regard to Judaism the same trend is found in the Hartwick as in the Harvard-Radcliffe sample, a shift away from the faith of the upbringing, although in a far less pronounced fashion: Hartwick male, 8 to 6 per cent as against Harvard 17 to 6 and Radcliffe 18 to 8 per cent. The Hartwick sample did not include any women of the Jewish faith.

e. There is a significant shift toward some new or individual type of religious belief in the Hartwick as in the Harvard-Radcliffe sample, but again to a somewhat lesser extent, 7 per cent for the entire Hartwick sample as against 11 per cent for Harvard and Radcliffe.

f. The Hartwick findings are in full agreement with the Allport conclusion from the Harvard-Radcliffe inventory that "while the more orthodox positions hold their members relatively well, they attract virtually no converts (2)."

g. Professor Allport reported that in the Harvard-Radcliffe investigation "the shift is invariably toward a more 'liberal' position, or out of the field of religion altogether (2)." The findings of the Hartwick project agree with this trend, if not invariably at least in by far the great majority of cases, including the "out of religion altogether" category, 14 per cent of the entire Hartwick sample. Of this 14 per cent all reported a religious, almost entirely Protestant, orientation in their upbringing.

h. When the various categories of shift from the religion of their upbringing are combined, we see, in the following subtable, the picture for the Hartwick sample as compared with the Harvard-Radcliffe total. Unfortunately there was no breakdown of the total Harvard-Radcliffe percentage, and therefore no comparison of the Hartwick male with Harvard nor of the Hartwick female with Radcliffe was possible.

SUBTABLE

COMPARISON OF SHIFT AWAY FROM FAITH OF UPBRINGING[5]

(In Percentages)

	Hartwick male	Hartwick female	Hartwick total	Harvard-Radcliffe total
Shift, i.e., turned against the faith of their fathers	54	37	47	58
Stayed within the faith of their fathers	46	63	53	42

Professor Allport said of the Harvard-Radcliffe "staying" figure "We are left with precisely 42 per cent who are content to stay within the tradition in which they were reared (2, p. 14)." His general conclusion in regard to this phenomenon was that "Fully half of our students, then, have turned against the faith of their fathers (2)." For the Hartwick sample as a whole almost the same statement can be made, "Just short of half" would replace "Fully half." Of the Hartwick men, exactly the same statement can be made: "Fully half, or, still more exactly, slightly more

---

[5] The Hartwick combined percentages of shift away from the faith of their fathers was obtained as follows:

In Percentages

	Hartwick (male)	Hartwick (female)	Hartwick (total)
Ethical	5	4	4
New	8	4	7
Non-religious	19	4	14
Protestant shift	22	25	22
Total shift from the faith of their fathers	54	37	47
Remained in the faith of their fathers	46	63	53

than half, of the Hartwick men "have turned against the faith of their fathers." Of the Hartwick women somewhat over one third have turned against the faith of their fathers.

Our general conclusion concerning the comparison between Hartwick and Harvard-Radcliffe students in regard to the question of remaining within the faith of their upbringing is that there is little significant, and no decisive, difference between the small denominational college and the large secular university. Not only is the trend the same, but many of the comparative percentage figures are quite close.

4. *Comparative Correlations of the Student's Faith with His Parents', at Hartwick, Harvard and Radcliffe*

The first correlation of the student's faith with the faith of his parents is the subject of Question 8 of the questionnaire. Question 8 asks the student to compare the firmness of his belief in religion with the firmness of his mother's and father's beliefs. The findings at Hartwick, Harvard and Radcliffe are presented in Table 5.

TABLE 5
STUDENT'S COMPARISON OF OWN AND PARENTS' FAITH AT HARTWICK, HARVARD AND RADCLIFFE, IN PERCENTAGES

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard (total)	Radcliffe (total)
Mother's					
More firm	19	17	18	7	14
Less firm	26	21	25	51	41
About the same	23	42	29	32	27
Don't know	11	0	8	10	18
Different	8	20	11	0	0
Blank	13	0	9	0	0
Father's					
More firm	23	21	22	16	23
Less firm	19	17	20	27	27
About the same	23	33	25	36	24
Don't know	13	21	15	21	26
Different	9	8	9	0	0
Blank	13	0	9	0	0

Professor Allport drew four conclusions from the Harvard-Radcliffe findings. We will compare our results with his conclusions.

a. Professor Allport drew as a first conclusion from the Harvard-Radcliffe sample "That students seldom regard themselves as more religious than they believe their parents to be (2, p. 15)." This holds generally for the Hartwick sample, but a higher percentage than Harvard reports a "more firm" faith, while the Hartwick-Radcliffe comparative percentages are more nearly equal. At least the tendency is the same in the two sets of samples.

b. Professor Allport drew as a second conclusion from the Harvard-Radcliffe inventory that "They (the students) more often consider their mothers' faith to be stronger than their fathers'." This also holds for the Hartwick sample but less markedly.

c. Professor Allport drew as a third conclusion that the Harvard-Radcliffe students "are more ignorant of their fathers' views on religion than of their mothers'," and he adds that this finding "reflects the practice in our culture for mothers to be the mentors of idealism within the family structure." The conclusion fully holds for the Hartwick sample and thus reinforces the Allport thesis concerning American mothers being the mentors of idealism, and thereby presumably of religion, within the family.

d. Professor Allport drew as a fourth conclusion, from the Harvard-Radcliffe sample, that "the evidence of Table 5 like that of Table 4 (his tables 7 and 6) clearly indicates a loosening of religious ties in the younger generation as compared with the older." This conclusion is based on the "Less firm" category in Table 5 in which Harvard and Radcliffe show high percentages particularly in relation to the mothers' beliefs. The "less firm" category for Hartwick students is lower by almost one-half than the Harvard-Radcliffe, again particularly in relation to the mothers' belief. This would indicate that while the trend toward loosening of religious ties is present in the Hartwick sample, it is less pronounced than in the Harvard-Radcliffe inventory. In regard to the comparison with the belief of fathers, however, the difference is far less, a fact which strengthens the evidence for the Hartwick trend toward loosening of religious ties in the younger generation. Here we conclude again that the trend at Hartwick is similar to that at Harvard and Radcliffe but in this instance to a significantly less degree.

Professor Allport draws an important general conclusion from the Harvard-Radcliffe samples in the last two tables: "Finally, and without the possibility of doubt, the results in Tables 4 and 5 (his Tables 6 and 7) do reflect a genuine trend toward secularization (2, p. 16)." Since we have already concluded that the tendencies found at Harvard and Radcliffe in these two tables are also reflected in the Hartwick findings, we now conclude that a similar secular trend exists at Hartwick. In view of the lesser degree of the tendencies at Hartwick, however, we at the same time conclude that the secular trend, though most certainly present, is less marked at the small denominational college than at the large secular institution.

The second correlation of the student's faith with the faith of his parents is the subject of Question 5. Question 5 asks the student to indicate whether there has been a period in which he reacted either partially or wholly against the beliefs of his parents. Table 6 presents the results of this question in correlation with the division of the positively oriented religious students into non-Catholic and Catholic.. Here Catholic is limited to Roman Catholic.

Table 6 shows that rebellion against the religion of parents on the part of non-Catholic and Catholics alike is if anything more common among Hartwick students, male, female and total, than among Harvard and Radcliffe students. In both the Hartwick and Harvard-Radcliffe inventories the table likewise indicates that rebellion against religious training is somewhat more common among Protestants and Jews than among Roman Catholics. The generally high percentage regardless of faith and the relatively slight difference between non-Catholics and Catholics would seem to indicate that such rebellion marks a more or less universal stage of development of modern American youth. The median age for such rebellion at both Hartwick and Harvard-Radcliffe is around 14.5 for girls and 15.5 for boys. In all respects, then, Hartwick conforms to the pattern of rebellion at some point against the religion of parents.

TABLE 6

PERCENTAGES OF STUDENTS REPORTING REACTION AGAINST RELIGION OF PARENTS, AT HARTWICK, HARVARD AND RADCLIFFE

"If you were brought up under some religious influence, has there been a period in which you have reacted either partially or wholly against the beliefs taught?"

	Hartwick male + Catholic	Hartwick male ++ Catholic	Hartwick female + Catholic	Hartwick female ++ Catholic	Hartwick total Catholic +	Hartwick total Catholic ++	Harvard + Catholic	Harvard ++ Catholic
Yes	89	75	80	75	83	77	73	62
No	11	18	15	25	17	17	18	28
Doubtful	0	7	5	0	0	6	9	10

+ Non-Catholic
++ Roman Catholic

5. *How Students View the Relationship Between Science and Religion at Hartwick and Harvard*[6]

In Question 12 of the questionnaire the students were asked to evaluate the relationship between science and religion in terms of support or degrees of conflict ranging from the negligible to irreconcilable. Table 7 presents the findings for the Hartwick and Harvard samples.

---

[6] Radcliffe was not included in the Allport Table.

From this table several interesting comparative results emerge.

a. Exactly the same percentage of the two samples feels that science and religion "clearly support one another." Here the small denominational college shows no difference whatever in comparison with the large secular university. Within the Hartwick sample itself there is no difference on this particular point between male and female students.

TABLE 7

COMPARATIVE STUDENT EVALUATION OF THE RELATION BETWEEN SCIENCE AND RELIGION AT HARTWICK AND AT HARVARD

(In Percentages)

	Hartwick male (N 53)	Hartwick female (N 24)	Hartwick total (N 77)	Harvard (N—386)
Religion and science clearly support one another	21	21	21	21
The conflict is negligible	24	29	26	32
The conflict is considerable but probably not irreconcilable	15	21	17	17
The conflict is considerable but not irreconcilable	17	16	17	14
Conflict is definitely irreconcilable	11	4	9	16
Special	12	9	10	0

b. A significantly lower percentage of the Hartwick than of the Harvard total sample reports that the conflict between science and religion is "negligible (more apparent than real)," 26 to 32. An even more markedly lower percentage of the Hartwick men, as compared with Harvard, feel that the conflict is negligible, 24 to 32. The Hartwick women show a lower percentage who feel that the conflict is negligible than Harvard, 29 to 32. In this category then, there is indeed a difference between the small denominational college and the secular university, but it is in the direction of more rather than less conflict between science and religion.

c. The Hartwick total and Harvard agree, percentagewise, that the "conflict is considerable but probably not irreconcilable." Here there is no difference between the two samples. A higher percentage of Hartwick women, however, select this category than do Hartwick men.

d. A higher percentage of Hartwick men, women and total than of

Harvard students feel that the conflict between science and religion is "considerable, perhaps irreconcilable," 17, 16, and 17 to 14. Here the denominational college reports more irreconcilable conflict than the secular university.

e. On the other hand, Harvard shows a higher percentage than does Hartwick in the category of "definitely irreconcilable," 16 to 9. The percentage of Hartwick men, however, is closer to the Harvard figure, 11 to 16.

f. By combining the first three categories in Table 7, we get the percentage of each sample who do not feel that science and religion are irreconcilable. The results are given in the following subtable:

SUBTABLE
PERCENTAGE WHO FEEL THAT CONFLICT BETWEEN SCIENCE AND RELIGION IS NOT IRRECONCILABLE

Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard
60	51	64	70

Thus there is a higher percentage at Harvard than at Hartwick who feel that the conflict between science and religion is not irreconcilable. Conversely a higher percentage at Harvard than at Hartwick feel that it is irreconcilable, 30 to 26 per cent.[7]

Any general conclusion about the comparative views on the relation between science and religion is difficult to draw. This is the case because, taken by itself, none of the categories of choice, except the "support" one, indicate from which point of view, science or religion, the choice was made. For example, if a student answers that the conflict is "irreconcilable" it could mean that he *accepts* the irreconcilability of the two and thereby retains his belief in religion. It does not necessarily mean that he favors science over religion. The presumption, from the fact that the Hartwick trend toward liberal beliefs, secularism and irreligion is generally less pronounced than the similar Harvard trend, would be that more Hartwick than Harvard students would *span* the "irreconcilability" of science and religion by means of religious beliefs. Be this as it may in the one category in which the answer can be definitive, namely "support one another," the denominational college and the secular university reveal absolutely no difference.

One very general conclusion can be drawn. From the comparison in Table 7 it is clear that there is at least as much awareness of felt conflict between religion and science, regardless of from which position, in the small denominational college as in the large secular university. It might be more accurate to conclude from the table that students of the denominational college feel the conflict more sharply. This may be accounted for at least in part by the constant contact with religion as well as science in the college curriculum.

---

[7] This is computed through a combination of the fourth and fifth categories.

6. *Students' Self Evaluation in Regard to Religious Beliefs at Hartwick, Harvard and Radcliffe*

The subject matter of this section is composed of two types of directly introspective self evaluation. First, the students compare their own beliefs to those of their contemporaries. Table 8 presents the results:

TABLE 8

"How Would You Say That Your Own Religious Sentiments and Needs Compare With Those of Your Contemporaries

(In Percentages)

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
Stronger than average	32	33	32	26	34
About average	42	58	47	40	34
Less strong than average	19	0	13	34	32
Blank	7	9	8	0	0

Here the two sets of samples are about equal in the "stronger than average" category, but Hartwick students cluster more firmly within the "about average" range and reveal a considerably smaller percentage in the "less strong than average" category. Since we do not know what kind of attitude, whether positively or negatively oriented toward religion, for example, little can be concluded from this comparative inventory. Whether it shows, as Professor Allport and his Associates claim, the excellence of the students' insight into their own beliefs (2, p. 19), is in our opinion indeterminable, since none of the terms are objectively referable. What is the average to which the student refers in his evaluation? And what are the religious sentiments and needs that are being compared?

In regard to the findings presented in Table 8, it can be concluded that in the last two categories there is a marked difference between the Hartwick and Harvard-Radcliffe samples, but what this difference signifies is impossible to determine. One thing seems certain, it is not a measure of the students' excellence of self-knowledge. Another is that it tells us nothing about the character or degree of the students' religious sentiments or attitudes.

The second self evaluation question is no more meaningful for purposes of comparison than the first. Question 10 of the questionnaire asks the student, "Do you feel that your views regarding religion, no matter what they are, in any way mark you off from your contemporaries, so that you sometimes feel embarrassed, isolated or discriminated against because of these views?" Table 9 presents the findings:

TABLE 9
"DO YOU FEEL THAT YOUR VIEWS ON RELIGION SET YOU APART, MAKE YOU FEEL EMBARRASSED OR ISOLATED?"
(IN PERCENTAGES)

	Hartwick male (N—53)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N—414)	Radcliffe (N—86)
Yes	34	37	35	17	28
No	66	63	65	73	59
Doubtful	0	0	0	10	13

On the one hand, the table shows that the Hartwick, Harvard and Radcliffe samples agree that some two-thirds of the students report that they have no self conscious awareness of being set apart, embarrassed or isolated because of their beliefs. On the other hand, the findings reveal that over twice as many Hartwick as Harvard or Radcliffe students report that they feel set apart, isolated or embarrassed by their religious views. No Hartwick students are doubtful in regard to their self awareness on this point, while 10 per cent of Harvard and 13 per cent of Radcliffe students report doubt in the matter. Thus there are differences between the Hartwick and Harvard-Radcliffe samples in the "yes" and "doubtful" categories. This, of course, tells us nothing about either the views, whether they are religious or irreligious, or about the particular "contemporaries" who presumably make the students aware of their "set apart," "isolated" or "embarrassed states of mind. Could it be that the atmosphere of the type of college might play a part in the student's self awareness of these states, if not in inducing the states themselves? Perhaps a large secular university furnishes a different atmosphere and therefore has a different effect in these respects than the small denominational college. It may be that the self evaluations and self awarenesses depicted in the two tables of this section indicate more about the environment, academic and otherwise, which are reflected in the self-consciousness of the individuals, than they do about the attitudes of the individuals involved in the Hartwick and Harvard-Radcliffe samples. Statistical analysis of student attitudes is the subject of this inventory. At any rate, such directly introspective reports basically reveal nothing significant about student attitudes and certainly nothing about the student environment which could not better be obtained by more objective methods. The presence

TABLE 10[8]

(Question 13)

COMPARISON OF STUDENTS' VIEWS CONCERNING THE ORGANIZED RELIGIOUS INSTITUTION (CHURCH, TEMPLE. ETC.) AT HARTWICK, HARVARD, RADCLIFFE AND WISCONSIN (1930)

(In Percentages)

	Hartwick male (N—53)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N—170)	Radcliffe (N—63)	Wisconsin —1930— (N—3010)
(1) The Church is the one sure and infallible foundation of civilized life	11	8	10	6	6	4
(2) On the whole the Church stands for the best in human life	42	71	51	36	40	24
(3) There is a certain doubt. Possible that Church may do a good deal of harm	13	13	13	18	13	38
(4) The total influence of the Church may be on the whole harmful	8	0	5	6	2	14
(5) The Church is the stronghold of much that is unwholesome and dangerous to human welfare	2	4	3	10	6	14
(6) Insufficient familiarity	4	0	3	4	8	6[9]
(7) A different attitude	20	4	15	20	25	

---

[8] Allport, Gillespie and Young took Question 13, the subject of Table 10, from a questionnaire used at Wisconsin in 1930, and included the results in their table 12, op. cit. p. 20.

[9] "Wisconsin students were urged to avoid the last two options on this scale. Harvard and Radcliffe students were freely allowed to elect them." (Allport, Gillespie and Young note) Hartwick students likewise were uninstructed and therefore free to elect them, H. K. W.

of religious prejudice and of prejudice against irreligion can be ascertained far better and more accurately by a direct examination of the student environment than by introspection.

The findings set forth in Tables 8 and 9 are not, in our opinion, concerned with statistical analysis of student attitudes and therefore are not really germaine to the investigation at hand. "Interesting but not relevant" might better express this position.

7. *What Students Believe and Do Not Believe at Hartwick, Harvard and Radcliffe*

In this section we present and analyze six tables, all concerned with doctrinal beliefs such as belief in the Church, in God, in the Person of Christ, and in Immortality. These are perhaps the most interesting of all the tables.

We begin with an inventory of attitudes toward the organized religious institution, which we here call generically "The Church."

Table 10 reveals that the Hartwick sample has a higher percentage of students than Harvard, Radcliffe and Wisconsin, who rate the church positively (first two options). There is less difference in the percentages of negative ratings given the church, (options (3), (4) and (5), except in the case of Wisconsin, particularly in option (3)). Table 10 indicates, however, that on the whole there is a significant difference between the small denominational college and the large secular universities in regard to views concerning the Church or the organized religious institution.

TABLE 11a[10]
(Question 14)
BELIEF IN GOD, AGNOSTICISM AND ATHEISM AT HARTWICK, HARVARD AND RADCLIFFE
(In Percentages)[11]

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
Belief in God	57	83	65	58	66
Agnosticism	17	13	16	21	12
Atheism	13	4	10	11	11

Table 11a reveals that approximately two-thirds of the Hartwick sample and slightly less than two-thirds of the Harvard-Radcliffe sample believe in a God of some type. What percentages of this belief are Theist, Deist or Pantheist are to be found in Table 11b. Slightly over three-tenths of both the Hartwick and Harvard-Radcliffe samples are Agnostics and approximately one-tenth of both samples are Atheists.

---

[10] Breakdown of Hartwick, Harvard and Radcliffe findings in optional answers to Question 14 of the questionnaire.

[11] Does not include "special" and "no opinion" categories and therefore per-cent does not add up to 100 per cent.

Thus there is little or no appreciable difference between the small denominational college and the large secular university in regard to belief in God, Agnosticism and Atheism.

Table 11b reveals that approximately one-third of the Hartwick sample against one-fourth of the Harvard-Radcliffe sample believe in a theistic God and therefore are orthodox Protestants, Catholics, or Jews. Slightly less than one-fourth of both the Hartwick and Harvard-Radcliffe samples are deist and therefore are unorthodox or liberal Protestants or Jews. Finally almost one-tenth of the Hartwick and Harvard-Radcliffe samples are pantheist and therefore are religiously, individual and radical, beyond the pale of both Christianity and Judaism.

TABLE 11b[12]
(Question 14)
(Optional Answers (1) (2) and (3)
TYPE OF BELIEF IN GOD (THEISTIC, DEISTIC, AND PANTHEISTIC) AT HARTWICK, HARVARD AND RADCLIFFE
(In Percentages)[13]

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
Theistic (Personal God)	32	38	34	19	40
Deistic (Impersonal God)	15	38	22	25	19
Pantheistic (God in Nature and Man)	9	8	9	11	7

Table 11(b) thus shows little or no difference in the beliefs of students in the small denominational college and the large secular university, with the single exception of the theistic (orthodox), category in which there is a difference of one-twelfth in favor of Hartwick theism.

Table 12 reveals that just over one-half of the Hartwick sample as against approximately one-third of the Harvard-Radcliffe sample believe that Christ is the Human Incarnation of God and thereby accept the central doctrine of

---

[12] Breakdown in the "Belief in God" category of Table 11a.

[13] Only first three optional answers to Question 14, so percentages do not add up to 100.

Catholic and Protestant Christianity. On the other hand, somewhere between one-fourth and one-third of the Hartwick sample reject this central doctrine of all Christianity by believing that Christ was not Divine but was a human being and a great teacher or prophet. Two per cent of both the Hartwick and Harvard samples believe that Christ is a mythical figure. It is interesting to note in this connection that twice as many Hartwick as Harvard men believe that Christ is a mythical figure, while no women of either Hartwick or Radcliffe so believe. With the exception of Radcliffe, all the samples agree on similar high percentages in the "none of these alternatives" category, a fact which signifies that there exists a large individualist, non-conformist trend in the beliefs of students.

TABLE 12

(Question 15)

BELIEFS OF STUDENTS CONCERNING THE PERSON OF CHRIST, AT HARTWICK, HARVARD AND RADCLIFFE (In Percentages)

	Hartwick male (N—53)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N—411)	Radcliffe (N 86)
(1) Christ is the Human Incarnation of God	43	71	53	25	42
(2) Christ was a great man, teacher or prophet	36	13	29	56	50
(3) Christ is probably a mythical figure	4	0	2	2	0
(4) None of these alternatives	17	16	16	17	8

We conclude that in regard to beliefs concerning the Person of Christ the small denominational college, while on the one hand being more orthodox, at the same time reveals, although to a somewhat lesser degree, the same trend as Harvard and Radcliffe, namely, away from orthodoxy toward liberalism, secularism, irreligion, and individual (non-conformist) thinking. In the Hartwick sample, concerning the Person of Christ, 47 per cent (a combination of options (2), (3), and (4)) fall within this general trend. On this pivotal doctrinal point, the small denominational college is split almost down the middle.

TABLE 13a[14]
(Question 16)
BELIEF AND DISBELIEF IN PERSONAL IMMORTALITY AT HARTWICK, HARVARD AND RADCLIFFE (In Percentages)

	Hartwick male (N—53)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N—411)	Radcliffe (N 85)
Believe in personal immortality	42	57	49	25	35
Disbelieve in personal immortality	58	43	51	75	65

Table 13a shows that roughly one-half of the Hartwick sample and less than one-third of the Harvard-Radcliffe sample believe in Personal Immortality, which keeps them within the bounds of Catholic and Protestant Christianity and of Judaism on this doctrinal point. On the other hand, fully one-half of the Hartwick sample and between two-thirds and three-fourths of the Harvard-Radcliffe sample disbelieve in Personal Immortality and thereby place themselves outside the bounds of Catholic and Protestant Christianity and of Judaism on this doctrinal point.

Here again we see that the small denominational college follows the same trend, though to a lesser degree, that Harvard and Radcliffe revealed in the original Allport, Gillespie, and Young Report toward liberalism, secularism,

TABLE 13b[15]
(Question 16)
(Options (1) —and— (2))
TYPES OF BELIEF IN PERSONAL IMMORTALITY AT HARTWICK, HARVARD AND RADCLIFFE (IN PERCENTAGES)[16]

	Hartwick (male)	Hartwick (female)	Hartwick (total)	Harvard	Radcliffe
Belief in immortality of the soul	40	53	46	23	34
Belief in Reincarnation	2	4	3	2	1

---

[14] This table breaks down Question 16 into two categories: Belief in personal immortality, being options (1) and (2); and disbelief in personal immortality being options (3), (4), (5) and (6).

[15] Breakdown of the Belief in Personal Immortality into the two categories: Immortality of Soul, and Reincarnation, options (1) and (2) of Question 16.

[16] Since this is a partial breakdown of Question 16, the percentages do not add up to 100.

individualism in the form of nonconformity, and, at the extreme, out of religion altogether. The difference between the two types of colleges is thus once again *shown* to be a matter of degree not of kind, and not too wide a difference of degree either.

By breaking down the Belief in Personal Immortality into Immortality of the Soul and Reincarnation, Table 13b shows that only 46 per cent of the Hartwick sample believe in the Judaic-Christian doctrine of the immortal soul leaving 54 per cent beyond the bounds of the Judaic-Christian faith in this respect. A more exaggerated version of the same general phenomenon is represented in the Harvard percentage of students, 23 and the Radcliffe percentage, 34, who stay within the fold of the Judaic-Christian doctrine of the immortal soul. The converse percentages for Harvard and Radcliffe, 77 and 66 respectively, present a startling picture of student non-conformity to the Judaic-Christian doctrine of the personally immortal soul. From Hartwick's 54 per cent non-conformity to Radcliffe's 66 and Harvard's 77 is a progression in degree, all constituting degrees of over-half non-conformity to the Judaic-Christian doctrine of the immortal soul.

The general conclusion from all the doctrinal tables in this section of the report, with the possible exception of Table 10, on the Church, presents substantial evidence which buttresses the developing hypothesis, namely, that the small denominational college follows the trend of the large secular colleges in religious beliefs and attitudes, toward liberalism, non-conformity with tradition, individualism, secularism, and on to irreligion. It lags behind, sometimes more sometimes less, but generally it is traveling along the very same road—at least according to the great majority of the tables so far presented and analyzed.

8. *Religious Practices of Students at Hartwick, Harvard and Radcliffe*

The religious practices of students is the subject of Questions 11a, 11b, and 11c of the Questionnaire. These questions are concerned with the frequency of Church attendance, prayer and experienced feeling of reverence. The findings are presented in Table 14.

In general the percentage of the Hartwick sample in each category falls in between the percentages of the Harvard and Radcliffe samples, sometimes closer to Harvard but more often closer to Radcliffe. Invariably the Hartwick female percentages run considerably higher than those of the Hartwick male in the religiously positive categories and lower in the religiously negative options. While the Hartwick male percentages more closely approach those of Harvard in both the religiously positive and negative categories, the Hartwick female percentages in the religiously positive options consistently exceed those of Radcliffe and in the religiously negative categories are markedly lower.

We see that one-third of the Hartwick sample report weekly attendance at church, as compared with Harvard, less than one-fifth, and Radcliffe, more than one-third. One-fifth of the Hartwick sample reports attendance at church only once or twice in six months, as compared to one-third of the Harvard and one-fourth of the Radcliffe samples.

TABLE 14
PERCENTAGES OF STUDENTS REPORTING VARIOUS RELIGIOUS PRACTICES DURING PREVIOUS SIX MONTHS, AT HARTWICK, HARVARD AND RADCLIFFE

	Hartwick male (N—53)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N 413)	Radcliffe (N 86)
		Attended Church			
About once a week	26	50	34	17	39
Every other week	13	17	14	7	9
Once a month	14	9	13	14	10
Once or twice only	26	13	22	32	26
Not at all	21	11	17	30	16
		Prayed			
Daily	14	46	25	15	35
Fairly frequently	14	22	16	11	18
Occasionally	22	22	22	15	8
Rarely	20	10	17	23	12
Never	30	0	20	36	27
		Experienced feeling of reverence			
Daily	13	25	17	9	17
Frequently	23	38	27	13	23
Occasionally	30	21	27	21	29
Rarely	14	8	13	19	9
Never	20	8	16	38	22

In regard to frequency of praying, one-fourth of the Hartwick sample reports daily praying as compared with approximately one-sixth of the Harvard and one-third of the Radcliffe samples. One-sixth of the Hartwick sample reports praying rarely in six months as against one-fifth of the Harvard and one-eighth of the Radcliffe sample.

Table 14 shows that one-sixth of the Hartwick sample expericenced daily feelings of reverence in the past six months compared to one-tenth of the Harvard and slightly over one-sixth of the Radcliffe samples. On the other hand one-eighth of the Hartwick sample reports feeling reverence rarely in six months as against one-twentieth of the Harvard and one-tenth of the Radcliffe samples.

Most revealing, however, is Table 14a presenting a breakdown of Table 14 into two categories: The first commbining all those options reporting some frequency of religious practices during the previous six months; the second composed only of that option reporting no religious practices whatever during the previous six months.

When we compare the percentages of religious practices presented in Table 14a with the percentages of doctrinal Christian (including Catholic

TABLE 14a

PERCENTAGE OF STUDENTS AT HARTWICK, HARVARD AND RADCLIFFE REPORTING *Religious* AND *No Religious* PRACTICES DURING THE PREVIOUS SIX MONTH PERIOD

	Hartwick male (N—53)	Hartwick female (N—24)	Hartwick total (N—77)	Harvard (N—413)	Radcliffe (N—86)
		Church attendance			
Some frequency of church attendance	79	89	83	70	84
No church attendance	21	11	17	30	16
		Prayer			
Some frequency of praying	70	100	80	64	73
No praying	30	0	20	36	27
		Feeling of reverence			
Some frequency of feeling of reverence	80	92	84	62	78
No feeling of reverence	20	8	16	38	22

TABLE 14b

COMPARISON OF RELIGIOUS BELIEFS AND RELIGIOUS PRACTICES AT HARTWICK, HARVARD AND RADCLIFFE

(In Percentages)

	Hartwick	Harvard	Radcliffe
	Religious beliefs		
Theistic belief (Personal God)	34	19	40
Divinity of Christ	53	25	42
Personal immortality	46	23	34
	Religious practices[17]		
Some frequency of church attendance	83	70	84
Some frequency of praying	80	64	73
Some frequency of feeling of reverence	84	62	78

---

[17] During previous six months.

and Protestant) and Judaic beliefs presented in Tables 11a and 11b (On Belief in God), 12 (On the Person of Christ) and 13a and 13b (On Immortality), and their subtables, we find that there are considerably higher percentages for religious practices than for religious beliefs in the three sets of samples. Table 14b presents this comparison.

Here we see that the percentage of religious practices in each category and for each of the three samples is higher than the percentage of religious belief in each category and for each of the three samples by 1.5, 2, 3 and even 4 times.

A significant comparison among the Hartwick, Harvard and Radcliffe samples is obtained by counterposing the average percentage of beliefs and practices in the three samples and the discrepancies in each sample, between these averages. Table 14c presents the results of this comparison.

TABLE 14c
DISCREPANCY BETWEEN RELIGIOUS PRACTICES AND RELIGIOUS BELIEFS AT HARTWICK, HARVARD AND RADCLIFFE
(In Percentages)

	Average	Discrepancy between religious practice and beliefs
Hartwick:		
Religious beliefs	44	38
Religious practices	82	
Harvard:		
Religious beliefs	22	40
Religious practices	62	
Radcliffe:		
Religious beliefs	37	38
Religious practices	75	

Two highly interesting facts emerge in Table 14c: (*a*) If we look at the averages for religious beliefs and practices we find that in the Hartwick sample the average percentage for religious practices is almost double that for religious beliefs. In the Harvard sample they are almost triple and in the Radcliffe sample they are fully double. There is, then, an exceedingly sharp contradiction between religious theory (belief) and religious practice at Hartwick, Harvard and Radcliffe. (*b*) The measure of the comparative sharpness or acuteness of this contradiction is the actual percentage discrepancy between the percentage of religious beliefs and that of religious practices.

Table 14c reveals that this percentage discrepancy is almost precisely identical for the Hartwick, Harvard and Radcliffe samples, 38, 40 and 38 respectively. These discrepancy per cents, taken separately, constitute indices of the quantity of *sharpness* or acuteness characteristic of the contradiction within each sample. Taken together they make possible a quantitative comparison of the quality called "sharpness" or "acuteness," characteristic of the internal contradictions in each of the three samples. Thus the percentage index of sharpness or acuteness of the internal contradiction between religious beliefs and practices in the Hartwick, Harvard and Radcliffe samples is shown, by the 38, 40, 38 indices, to be basically the same.

This identity among the three samples in regard to the sharpness or acuteness of the internal contradiction within each between Christian-Judaic beliefs and Christian-Judaic practices, in no way signifies that the religious level of the internal contradictions is the same at Hartwick, Harvard and Radcliffe. Only the comparative sharpness of their internal contradictions is shown, by means of the percentage index, to be essentially the same.

To discover the comparative religious *levels* of the Hartwick, Harvard and Radcliffe samples, we must look again at the average percentage figures, but this time only at the percentage figures for religious beliefs in the three samples as presented in Table 14c. We notice that there is considerable variation among the percentages in regard to religious beliefs. These differences in the percentage figures for religious beliefs are the measure of the differences in religious level of the internal contradictions between religious beliefs and religious practices within each sample. The percentage figure for religious beliefs is what makes the religious levels comparable among the three samples. This is the case because the percentage of religious faith is the index of the religious level of the internal contradiction between belief and practice characteristic of the particular sample. The fact that this quality has a quantitative value in the form of the percentage figure of religious faith (originally derived by taking the average of the percentage figures for the various types of Christian-Judaic beliefs) makes the qualitative (religious or more specifically Christian-Judaic) levels of the three samples fully comparable.

Thus in the Hartwick sample, the religious *level* of the contradiction between Christian-Judaic beliefs and Christian-Judaic practices is given by the actual quantitative percentage figure of religious beliefs, namely 44. This percentage figure 44 thus constitutes the quantitative index for the qualitative (religious) level of the particular contradiction at Hartwick between religious beliefs and religious practices. This Hartwick contradiction is represented quantitatively by the percentage ratio 44/82. (Table 14c).

In the Harvard sample, the qualitative (religious or Christian-Judaic) level of the contradiction between religious beliefs and religious practices is given by the actual quantitative percentage figure for religious beliefs, namely 22, (see Table 14c). The percentage, 22, is thus the quantitative index for the qualitative (religious) level of the particular contradiction at Harvard between religious beliefs and religious practices. The given Harvard contradiction is represented quantitatively by the percentage ratio 22/62 (Table 14c).

In the Radcliffe sample, the quantitative index representing the qualitative (religious) level of its particular internal contradiction between religious beliefs and religious practices is 37 (Table 14c). The quantitative representation of the qualitatively particular Radcliffe internal contradiction between beliefs and practices is the percentage ratio 37/75 (Table 14c).

The fact that the *qualitative* (religious) *level* of all three samples is in each case represented quantitatively by their respective percentage index now makes it possible to compare Hartwick, Harvard and Radcliffe in regard to the *internal qualitative* (religious or Christian-Judaic) *level* of the contradiction between beliefs and practices characteristic of each. The qualitative comparison is made quantitatively by comparing the Hartwick level-of-contradiction index, 44, with the corresponding indices for Harvard and Radcliffe, 22 and 37 respectively.

Comparing first Hartwick with Harvard by means of the quantitative indices of the religious level of the internal contradictions between religious beliefs and practices, namely 44 and 22 inherent in each, we find that the level of the Hartwick sample is twice as religious as the level of the Harvard sample. The quantitative ratio furnishing a qualitative comparison can be expressed either to compare the religious level of Hartwick with Harvard, 44/22 or 2 to 1 (2/1) or Harvard's religious quality or level with Hartwick's, 22/44 or 1 to 2 (1/2).

Comparing Hartwick with Radcliffe in regard to the qualitative or religious levels of the two samples, we get the ratio 44/37 or 5/4, or Radcliffe with Hartwick, 37/44 or 4/5. Thus Hartwick is one-fourth more religious than Radcliffe and Radcliffe is four-fifths as religious as Hartwick.

The quantitative comparison of the religious level of Harvard and Radcliffe, in which we are really comparing the qualitative *levels* of their respective contradictions between religious beliefs and religious practices by means of the quantitative index of their religious beliefs, is 22/37 or 37/22.

In comparing Harvard to Radcliffe we have 22/37 or 3/5 which means that the religious level of the Harvard sample is three-fifths that of the Radcliffe sample. And, vice versa, that the level of the Radcliffe sample is 5/3 or two-fifths more religious than that of Harvard.[18]

D. Conclusion

Now if in combination we employ the two sets of indices, the percentage indices of the religious levels of the three samples and the percentage indices of the sharpness or acuteness of the internal contradictions between religious beliefs and religious practices within each college sample, we can come to a general conclusion in regard to the comparison of religious attitudes at Hartwick, Harvard and Radcliffe.

It will be recalled that the Hartwick inventory, with its tabulation and analysis in this report, was undertaken primarily to test the hypothesis advanced by Allport, Gillespie and Young to the effect that any investigation, such as that made by them at Harvard and Radcliffe, of a small denominational college would "almost certainly differ" in regard to findings. We are now ready to draw conclusions about the validity of the hypothesis on the basis of the accumulated weight of our findings, brought into focus by use of the two sets of indices.

---

[18] This treatment of the subject in terms of quantitative indices of qualitative levels is an example of a method by which a supposedly immeasurable quality can be expressed quantitatively and can render two supposedly incomparable qualities quantitatively comparable.

The method is based on the logical principle of the essential inseparable unity of quantity and quality in which there is quality within every quantity and vice versa. Every quantity expresses a quality. This principle makes it possible to see in mathematics a qualitative level within supposedly pure quantities, i.e., within numbers themselves. For example, all numbers fall into classes or sets which are in fact qualities or levels of numbers such as units, tens (or 'teens), twenties, etc., in the decimal system. In any number system the qualitative levels of supposedly "pure" quantities or numbers would be units, s, 2s, 3s, etc., and on to the s^2, s^3, etc., levels. Each level of numbers or quantities is a quality.

That numbers are "purely" quantitative is a total misconception. There are classes or qualities within numbers just as there are in everything else, nature or man-made, concrete or abstract. Mathematics has erroneously been considered to be so abstract as to be absolutely abstracted from quality and to be therefore exclusively quantitative. The fact is, on the contrary, that the more abstract quantity gets, the more abstract does quality become. It becomes so abstract as to be difficult to recognize as quality—as for example it is difficult to grasp the concept of qualitative levels of natural numbers such as twenties, thirties or hundreds. Once the concept is understood, the recognition of such quantitative qualities is very simple.

The misconception that mathematics is exclusively quantitative acts at the present time to retard the theoretical and practical development of mathematics itself as well as the applications of mathematics in those sciences that are supposed to be exclusively qualitative such as the psychology of religion. The misconception limits mathematics to statistics in these fields.

First, there is no difference whatever between the Hartwick, Harvard and Radcliffe samples in regard to the comparative sharpness or acuteness of the contradiction between religious, meaning Christian-Judaic, beliefs on the one hand and religious, Christian-Judaic practices on the other. This quantitatively determined qualitative fact indicates that the dominant situation in regard to all three samples is one of varying degrees of breaking away *in theory* or belief from the religious, Christian-Judaic, doctrines of the two major religious systems in America, but at the same time continuing the religious *practices* of going to church, praying and experiencing feelings of reverence. This contradiction is as true of the small denominational as of the large secular colleges. All the tables in this report support this conclusion, made explicit by the index of the sharpness and acuteness of the contradiction between religious beliefs and religious practices.

There is no difference among the samples in regard to either the *existence* of this contradiction or its *degree* of sharpness. There is, however, a considerable difference in the qualitative or religious level of the contradiction, as measured by the relevant indices. The denominational college has a considerably higher religious level of this contradiction, twice as high as Harvard and one-fourth higher than Radcliffe. What does this mean? In light of the contradiction between religious theory and religious practice, common to both the denominational and secular colleges, it can only mean that, while students at the denominational schools are moving away from traditional religious doctrines, they have not moved as far away as have their fellow students in the large secular universities.

The Hartwick inventory in the most essential respects is not in agreement with the Allport, Gillespie and Young hypothesis in regard to expectations of difference between the religious attitudes of students at a small denominational college and those of students at a large secular university. At the same time, there is a difference between the two. This difference is the less essential one of *degree* of movement away from doctrinal beliefs with a concommitant sharp disparity in the correlation of beliefs with practices. The *identity* between the two types of institutions of higher learning is one of *kind,* while the difference is one of *degree.* Thus in the less essential aspect, the Hartwick inventory supports the Allport, Gillespie and Young hypothesis.

By and large there is, among Hartwick men, a striking degree of identity in both the trend and the amount of movement along the secular direction of the trend. By and large the Hartwick male sample falls somewhere between that of Harvard and Radcliffe, more often than not closer to the former than the latter, and therefore there is very little difference of degree and of course none of kind.

The Hartwick female sample, on the other hand, while following the general trend in a discernible fashion, does so with such a difference, falling considerably beyond the religious level of Radcliffe, that it belongs for the most part almost in a class by itself. Due to the fact, however, of the high and fully commensurate index of sharpness of conflict between religious beliefs and religious practices, we have to conclude that the Hartwick female sample falls, somewhat precariously, within the general trend and therefore can be classed as being within the identity of kind with difference being a matter of degree, but a high degree. The other three samples are by and large closer to each other while the Hartwick women stand apart, members of the class but hovering close to the periphery.

No attempt can here be made to account for the findings at Hartwick. This would require a sociological and perhaps a historical treatise. Anything that could be said would be too general, abstract and perhaps speculative.

REFERENCES

1. ALLPORT, G. W. The Individual and His Religion. New York: 1950.
2. ALLPORT, G. W., GILLESPIE, J. M., & YOUNG, J. The religion of the post-war college student. *J. Psychol.*, 1948, **25**, 3-33.

Department of Psychology
Hartwick College
Oneonta, New York

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 383-385.

HYPNO INTROSPECTION: A CONTRIBUTION TO THE THEORY OF HYPNOTHERAPY. PART I*

Research Division, Central Islip State Hospital, New York

SEYMOUR HALPERN

A. PURPOSE

The value of hypnotherapy while well established has inherent limitations. The considerable number of patients who are resistant to hypnotic induction markedly limit its applicability (6). Work from this laboratory has been directed towards elucidating the theoretical concepts as well as improving the techniques of hypnotherapy (1, 2, 3, 4, 5). The present report describes a further development and is designated the hypno-introspective approach.

B. THE HYPNO-INTROSPECTIVE APPROACH

The patient is requested to lie down on the couch. He is to assume a comfortable position, arms and legs extended and eyes closed. It is advisable to have patients remove their shoes. After settling, he is asked to remain motionless and to resist all impulses to stir. When this has been accomplished, the patient is asked to direct his attention to his bodily feelings; he is to report any sensations, affects, thoughts or images. When a bodily sensation is reported, he is asked to focus upon it until a change occurs.

If a patient reports unpleasant or painful sensations, he is asked to attend to the sensation and told "to let it build up"; e.g., when a patient states, "I am getting a headache," the therapist encourages him to experience the sensations fully. He is reassured that the pain is transient, represents a bodily communication and will reveal the nature of a particular conflict—that in the process of discharging the sensation he will achieve an equilibrium at a higher level of integration.

Frequently it is necessary to direct the patient's attention to observable tonic changes; these changes are often the only clues available for inferring alterations of psychic activity. Close observation of the patient is particularly important during intervals of silence. Patients are frequently unaware or do not report all changes. Commenting on tonic changes is helpful during periods when patients report an absence of bodily feelings. Their conflicts may

---

* Received in the Editorial Office on January 8, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

be posturally expressed—calling attention to a stiff neck, a grimace or an arched back, etc., aids the patient in reestablishing the associative process. If feelings of emptiness are reported, the patient is told to accept them and experience them fully; this invariably leads to changes in bodily feelings. Patients who express a strong desire to open their eyes are admonished to keep their eyes closed and to discuss their feelings.

All interventions are obtrusive by nature and care must be exercised in order to avoid interfering with the associative process. Not all patients can provide an adequate verbal commentary on the course of feelings and sensations and for some describing the introspective events results in the abolition of perceptualizations. For these patients long periods of silence may be frequent occurrences.

Initial investigations reveal that verbalizations of the introspective process may be unnecessary with young children. Young children are merely told that they will have feelings and perhaps pictures which will be discussed afterwards. In these situations, the task of the therapist is to remind the child not to move. A discussion of the experiences is conducted in the waking state after hypno-introspection.

During extended periods of silence the therapist's interventions are usually non-directive, e.g., "What's happening now." The frequency and quality of interpretive interventions are as in other analytic therapies. When the associations suggest a loss of contact with the body, the patient's attention might at the discretion of the therapist be redirected to his body. It should be noted that hypno-introspection stresses the feeling and affective aspects of consciousness. It is aimed at obviating the difficulties engendered by verbal communication. It seeks to accomplish this goal by confronting the patient with his immediate experiences. The patient is continually impressed with the need for recognizing and accepting the introspective events, particularly his feelings. The patient's attention is directed to all involuntary movements (twitches, trembling) which he is instructed to focus upon. It is often necessary to remind him to stop voluntary movements.

During the initial phases of therapy, failure to report changes in bodily feelings may signify a lack of understanding of a patient's therapeutic responsibilities. Such failures may also indicate resistance.

It might be pointed out that the patient's verbal report of his experiences during hypno-introspection only represents an approximation. The therapist's appreciation of the patient's reports will be enhanced when he has attempted similar experiences (1, 2). From the foregoing it is obvious that individual differences among patients requires therapeutic flexibility.

Proceeding in this way, the patient learns, at his own pace, to achieve deeper hypnotic states. As his capacity for achieving deeper hypnotic states increases, the advantages of hypnosis as a means of achieving self-understanding become increasingly possible.

C. CONCLUSION

The hypno-introspective approach has been used at our facilities during the past six months. It appears to offer significant advantages over the usual hypnotherapeutic methods. Further experiences will be needed to ascertain its full potential value. Preliminary observations suggest that it may be possible through this approach to extend the application of hypnotherapy to a wider range of clinical and experimental problems.

REFERENCES

1. HALPERN, S. The role of hypnosis in self understanding. *J. Psychol.,* 1961, **52**, 265-267.
2. ———. On the similarity between hypnotic and mescaline hallucinations. *Internatl. J. Clin. & Exper. Hypnosis,* 1961, **9**, 139-149.
3. HALPERN, S., & MERLIS, S. Hypnosis and psychotropic agents: their interactions in mental patients. *Neuro-Psychopharmacology,* 1961, **2**, 258-262.
4. HALPERN, S., & WHITE, L. Changes in body image complex through hypnotically induced moods. *J. Amer. Assoc. Soc. Psychiat.,* 1960, **1**, 176-194.
5. ———. Experimental hypnotherapy with a cancer patient. To be published in the *J. Psychosomat. Dentistry & Medicine,* 1962.
6. KLINE, M. V. Sensory-imagery techniques in hypnotherapy: psychosomatic considerations. *Topic. Probl. Psychother.,* 1960, **3**, 161-173.

Research Division
Central Islip State Hospital
Central Islip, New York

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 387-397.

PERSONAL BELIEFS OF SCHIZOPHRENICS AND THEIR PARENTS*

South Oaks Psychiatric Hospital, New York[1]

LEONARD W. KRINSKY

A. INTRODUCTION

The parent-child relationship has been investigated in recent years with regard to the etiology of schizophrenia and reports have indicated that parental attitudes and beliefs are significantly related to the patient and his illness. Lacking, however, have been systematic data in which the relationships of mothers to their patient offspring have been contrasted with those of fathers of the same schizophrenic children.

This study was undertaken specifically to investigate the differences between fathers and mothers in their identification with the personal beliefs of their patient offspring. Further objectives were to examine the differences between these same parents with regard to the projection of their own beliefs and their awareness of the beliefs of their children. A correlated objective was to investigate these relationships as a function of the chronicity of the patient's illness.

Of various available response indices, those purported to measure personal beliefs were selected for comparing mother-patient and father-patient relationships. This afforded specific comparisons with previous studies of the same problem. Cass (5) contrasted the pattern of likes and dislikes of mothers and their delinquent children as did Dworin and Wyant (9) regarding similarities in political beliefs and Bassell (4) in views on mental illness. Also considered was the fact that an individual's personal beliefs influence his mode of inter-relating with others (6, 13, 15) and that fathers and mothers of schizophrenics are not likely to be exceptions.

Mothers have been the primary subjects for studies regarding parent-patient relationships inasmuch as investigators have attempted to identify special features of the so-called "mothering" relationship. An extensive review of the literature (5, 6, 7, 10, 11, 14, 15, 17, 21) revealed that the mother of the schizophrenic was most frequently described as "hostile, rejecting, overcontrolling, overprotective, guilty and identified with her sick

---

* Received in the Editorial Office on January 9, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Partially supported by South Oaks Research Foundation Inc.

child." Many investigators describe this constellation of negative characteristics as characterizing the "schizophrenogenic mother."

Most of these investigations appeared to have limitations in research design which left the results open to question. Although some studies dealt with sizable *N*s and seemed to contain a consistent methodology, the conclusions were based on what often appeared to be clinical impressions of the investigator unchecked by other judges. A detailed critique of this literature has been offered by Krinsky (12) and the identity of the schizophrenogenic mother questioned.

Few studies have been concerned with fathers of the mentally ill. The literature, though limited (1, 7, 10, 11, 14, 19, 20, 21), has described him as "weak, passive, less noxious, less interested in the patient and less inclined to give the patient opportunity for identification than the mother." Reichard and Tillman (18) described one case as an opposite type, one which they termed the "schizophrenogenic father." He was said to be "sadistic, domineering and overly rejecting, a man who was trying to cover his own failures in life, who was dependent on his mother, and who usually had a masochistic wife."

In these several studies, including Reichard and Tillman's, the father was never directly questioned or observed. Inferences regarding his role in the family constellation were based on the response of either the patient or the mother but not the father himself. Also, as we noted in the studies of the mother of the patient, conclusions were based largely on subjective kinds of data unchecked against other observers.

These concepts of the schizophrenogenic parent suggested that with suitable measurement mothers would project and identify more with the personal beliefs of their schizophrenic children but at the same time have relatively little actual awareness compared with fathers.

Psychiatric writings (8, 16) have indicated that the longer a patient is ill, the more serious his mental condition must be considered and the more grave the prognosis. Therefore it seemed reasonable that the intensity of the disorder would be correlated with the degree to which the patient's father and/or mother typified the schizophrenogenic parent. In terms of the specific hypotheses of the present investigation, the parent of the chronic schizophrenic would be less aware, project more and identify less with the personal beliefs of the patient than parents of acutely ill offspring. Also possible was that chronicity would exaggerate differences between mothers and fathers.

B. METHOD

1. *Questionnaires*

Three questionnaires were employed in this study, namely those of Cass (5), Dworin and Wyant (9) and Bassell (4). By utilizing these scales it was possible to tap beliefs relating to relatively innocuous likes and dislikes, beliefs relating to societal structure, patriotism and social issues, and beliefs relating to mental illness.

Cass's questionnaire dealt with innocuous topics and consisted of eleven checklists, each checklist having on the average twenty items. *S* selected from among the items those three which were most preferred. For example: as answers to a lead question, one checklist required *S* to specify his three most preferred vocations, another his three major fears and a third, the three objects most preferred as personal possessions. The Dworin and Wyant questionnaire comprised sixteen items originally contained in the F Scale of Authoritarianism (2). Dworin and Wyant found these items differentiated significantly between mothers of schizophrenics and mothers of normals. These sixteen questions related to societal structure, patriotism and social issues. The third questionnaire related to beliefs concerning mental illness and contained only those twenty-six items which were found by Bassell to differentiate significantly between patients who were ready and those not ready for release from a hospital. These items referred to stereotypes, for example, "A person who is mentally ill has an evil spirit in him."

2. *Population Sample*

The *S*s for this study were 80 mothers and fathers and their patient offspring, the latter residents of a large metropolitan state hospital. All patients were males and were diagnosed as schizophrenic according to the usual diagnostic procedure employed in this hospital.

The selection procedure required that patients have both parents living and available for testing. This usually meant that they lived in the nearby metropolitan vicinity.

Of the patient group, 60 were seen within one month after their hospitalization. Of these 60, 40 were judged acutely ill by the psychiatrist in charge. The remaining 20 were judged "chronic." An additional 20 patients were those who had been in the hospital from one to three years and were then in a chronic building. These two samples of 20 each were combined to make up the sample labelled as "chronic." The results of a median test on the several measures employed in the investigation indicated that these two groups were homogeneous and could be treated as comprising a single sample.

The sample was also compared as a function of the number of hospitalizations for each patient. As seen in Figure 1, 52 patients and their parents comprise the first admission group and the remaining 28 patients and their parents make up the multiple admission group.

	Number of parents of acute	Number of parents of chronic
Father	40	40
Mother	40	40
	Number of parents of F. A.	Number of parents of M. A.
Father	52	28
Mother	52	28

FIGURE 1

DESIGN OF THE STUDY WITH NS FOR EACH OF THE SELECTED SAMPLES

All patients were male, ranged in age from 15 to 35 and averaged 24 years. The parents of these patients were between 38 and 65 years of age and averaged 54 for fathers and 51 for mothers.

Parents were "natural parents" who had lived with the patient through his first eight or more years. None was or had been patients under psychiatric care. As far as could be determined from patient's case histories, no parent was a known alcoholic or drug addict.

3. *Procedure*

Patients were administered the Cass, Dworin and Wyant, and Bassell questionnaires respectively. The parents of these patients were then tested on their next visit to the hospital. Usually this occurred before they saw the patient but in a few instances they were seen after their visit. If both parents arrived together they were tested in a large room at separate tables and were offered no opportunity to communicate. Otherwise they were tested at different times in the same room.

The same questionnaires that had been administered to the patients were administered twice to their parents. In one administration they were asked for their own beliefs and in the other, they were asked to check the responses that they believed their children would give if asked the same questions. The instructions were counterbalanced so that one half the sample gave their own responses first, the other half the responses they attributed to their child.

4. *Scoring*

The dependent measures were termed identification, awareness and projection and were derived by determining the number of identical responses given by parents and patients to the several "belief" questionnaires. These measures were originally devised by Cass. Identification was measured by counting the number of expressions of belief given by each parent for himself and which were also given by his or her son. Awareness was measured by the total number of item correspondence between a patient's responses and the responses which his parent attributed to him. Amount of projection was indicated by the degree of exact correspondence between a parent's responses and the responses which his parent attributed to the patient but which were not actually expressed by the patient himself.

These scores were in brief as follows:

Identification: Parent about self compared with patient about self.

Awareness: Parent about child compared with patient about self.

Projection: Parent about self compared with parent about child (measured if child's responses were not the same).

In the remainder of this study the scores Id I, Aw I, Proj I, refer to the scores derived from Cass's questionnaire; Id II, Aw II, Proj II refer to the scores derived from Dworin and Wyant's questionnaire; Id III, Aw III, Proj III, to the results of Bassell's questionnaire.

C. RESULTS

When the data were collected, values were computed for each of the dependent measures on each of the three short answer questionnaires. Thus each patient had three identification, three awareness and three projection scores.

Inasmuch as the data were discrete and asymmetrical with a clearly defined zero point and operationally defined upper limit for values obtained from each measuring instrument, the median test was employed. Each patient's scores were assembled in the form of a frequency distribution, the median was then computed, two by two contingency tables were constructed and a chi square value was obtained.

1. *Comparison of Mothers and Fathers*

Significant differences in parent-patient relationships were revealed only in connection with Cass's questionnaire which dealt with innocuous matters. Fathers ($N = 80$) had significantly higher identification (.05 level of confidence) and projection (.01 level of confidence) scores whereas mothers

($N = 80$) exhibited greater awareness (.05 level of confidence) with respect to beliefs held by patients.

These differences disappeared when groups of mothers and fathers were subdivided in terms of the acuteness and the chronicity of the patient's symptoms. However, when patients were reclassified as either first or multiple in their hospital admissions, differences between parent groups were again uncovered with respect to the Cass questionnaire. Fathers of multiple admission patients ($N = 28$) displayed greater identification and projection (.01 level of confidence) than mothers of the same patients ($N = 28$). This outcome did not occur in the case of first admission patients. The greater awareness of patient's beliefs was characteristic only of first admission mothers ($N = 52$) (.05 level of confidence). Thus differences in parent-patient relationships were in part a function of the number of mental hospital admissions of the patient and hypothesized differences could be sustained with this qualification.

2. *Comparison of Parents of Acute and Chronic Patients*

On the basis of chi square values obtained from median tests, the parents of acutely ill patients ($N = 80$) were more generally identified with and more aware of their son's beliefs than the parents of chronically ill patients ($N = 80$). Significant differences in parent-patient relationships were found on all three questionnaires, but results were most evident on the Bassell questionnaire, the one dealing with stereotypes of mental illness.

With respect to beliefs regarding mental illness the parent of the acutely ill patient identified more with the patient (.01 level of confidence), was more aware (.02 level of confidence) and projected less upon the patient (.01 level of confidence) than did the parent of the chronically ill patient. These results supported our hypothesis. On the basis of the Cass questionnaire (innocuous beliefs), parents of acutely ill patients again showed greater awareness (.01 level of confidence). A contrasting result was seen in relation to the Dworin and Wyant questionnaire (social issues etc.). Parents of acutely ill patients appeared to project more upon their sons than did the parents of the chronically ill patient (.02 level of confidence).

When mothers and fathers were considered separately, results indicated that mothers of acutely ill patients ($N = 40$) were generally more identified with, more aware of, and projected less with respect to the beliefs of patients than did mothers of chronically ill patients ($N = 40$). Such results as given in Table I were seen principally with regard to the Bassell questionnaire.

In a similar comparison, fathers of the acute patients ($N = 40$) were generally more identified (.05 level of confidence) and aware of the patient's beliefs (.05 level of confidence) than were the fathers of chronic patients ($N = 40$). This finding was made on the basis of the Cass questionnaire. With respect to projection of beliefs, this tendency was more evident for fathers of acutely ill patients on one scale (Dworin and Wyant) (.01 level of confidence) and less evident on another (Bassell) (.01 level of confidence).

TABLE 1
CHI SQUARE VALUES COMPARING FATHERS AND MOTHERS OF ACUTE AND CHRONIC PATIENTS

	Fathers: acute-chronic		Mothers: acute-chronic	
	Chi square	Level of confidence	Chi square	Level of confidence
Id I	5.00	.05	N. S.	
Id II	N. S.		N. S.	
Id III	N. S.		7.20	.01
Aw I	5.00	.05	7.20	.01
Aw II	N. S.		N. S.	
Aw III	N. S.		5.00	.05
Proj I	N. S.		N. S.	
Proj II	9.80	.01	N. S.	
Proj III	7.20	.01	5.00	.05

Table 2 summarizes the major results in terms of a set of generalizations. The hypotheses relating to differences between mothers and fathers were generally contradicted in terms of innocuous topics. The differences hypothesized between the chronic and acute groups extended to all questionnaires and were generally confirmed.

D. DISCUSSION

The results of this study which compared mothers and fathers had some correlation with observations by Arieti (3) but were generally contrary to the results of the majority of investigations. In contrast to the expected outcome of the mother of the schizophrenic being highly identified and lacking awareness of the patient, our group of mothers showed relatively low identification and high awareness in the area of personal beliefs. Also they were disinclined to project their beliefs onto their sons.

Differences between parents were consistently reliable in only the area of innocuous beliefs although results of earlier studies suggested that differences were generalized to a large number of beliefs. Admittedly the present finding could have been specific to the sample; nevertheless our evidence suggested that mother- and father-patient relationships have more similarities than differences. The differences which did appear yielded a description of fathers as projecting and identified with the beliefs of their patient-sons but lacking in awareness of what their sons actually thought on various items. The indifferent, withdrawn but understanding individual known as the schizophrenogenic father was not prevalent in our group. Rather than a passive member, fathers appeared to be an integral part of the family constellation and had more, rather than less, influence than mothers on the patient.

TABLE 2
SUMMARY OF THE MAJOR RESULTS AS THEY BEAR ON THE MAJOR HYPOTHESIS

Kind of sample comparison	Significant measures	Direction predicted	Direction obtained	*P*-value
		Mothers compared with fathers		
Total groups	Id I	mothers more	mothers less	(.05)
	Aw I	mothers less	mothers more	(.05)
	Proj I	mothers more	mothers less	(.05)
First admission	Aw I	mothers less	mothers more	(.05)
Multiple admission	Id I	mothers more	mothers less	(.01)
	Proj I	mothers more	mothers less	(.01)
		Acute compared with chronic		
Total groups	Id III	acute more	acute more	(.01)
	Aw I	acute more	acute more	(.01)
	Aw III	acute more	acute more	(.02)
	Proj II	acute less	acute more	(.02)
	Proj III	acute less	acute less	(.01)
Mothers	Id III	acute more	acute more	(.01)
	Aw I	acute more	acute more	(.01)
	Aw III	acute more	acute more	(.05)
	Proj III	acute less	acute less	(.05)
Fathers	Id I	acute more	acute more	(.05)
	Aw I	acute more	acute more	(.05)
	Proj II	acute less	acute more	(.01)
	Proj III	acute less	acute less	(.01)
		First admission compared with multiple admission		
Total groups	Aw II	F. A. more	F. A. more	(.05)
Fathers	Aw II	F. A. more	F. A. more	(.02)

The mothers' relatively greater awareness of their son's beliefs could be explained by the fact that they spent more time at home with the patient than did the father. This was particularly evident in the kinds of beliefs of which the mothers were aware, namely innocuous beliefs, were of a type readily observed. By merely being in the physical presence of the patient, a mother could have easily learned her son's vocational interests or preferences for personal possessions. The fact that this awareness did not carry over to supposedly more intense beliefs such as those measured by the scales on social issues and mental illness lent credence to this speculation. With regard to more intense beliefs, fathers exhibited as much or as little awareness of patient's beliefs as did mothers.

The available literature revealed no empirical studies of the differences between chronic and acute groups from the standpoint of parent-patient relationships and in this sense our findings were exploratory. The greater identification and awareness and the relative absence of projection in the case of parents of acute as contrasted with chronic patients suggested a number of interpretations. One possibility was that the chronic patient is more deviant, hence harder for his family to know. Consequently, his parents were unable to establish a close psychological affiliation with him especially in the area of personal beliefs. In the same sense, the differences between chronic and acute groups were conceivably the result of the illness, not the cause; that is, as a patient exhibited chronicity of symptoms, the beliefs of his family were altered to incorporate this fact. Equally plausible was that the greater the lack of identification and awareness of the parent, the farther psychologically the parent was from the patient and this, in turn, related to whether or not a patient became chronically ill.

Contrary to other scales, the scale dealing with social issues (Dworin and Wyant) revealed the parents of acute patients to project more rather than less as compared with parents of chronic patients. Projection was measured by the degree of correspondence between a parent's own beliefs and those which he attributed but which are not actually given by the child. A high projection score on the part of parents could therefore be understood to represent a resistance on the part of a child to parental ideas. This contention seemed pertinent to the "social issues" scale since many of the questions related to topics on which the "good" parent tries to instill stereotyped ideas in his child. Perhaps the acutely ill patient was more in contact with reality and therefore more critical and less ready than chronically ill patients to accept the stereotyped ideas of his parents.

E. Summary and Conclusions

Questionnaires designed to measure beliefs regarding innocuous topics, social issues and mental illness were administered once to 40 acutely and 40 chronically ill schizophrenic patients and twice to both their mothers and fathers. Parents indicated their own personal beliefs on one set of forms and recorded on the other set the beliefs which they thought belonged to their patient-offspring. By comparing parents' and patients' responses to various items, measures of the degree to which mothers and fathers identified, projected, and were actually aware of the personal beliefs of their sons were obtained.

The major findings were as follows:

1. Contrary to prediction, mothers and fathers appeared to be more alike than different in relating their beliefs to those of their patient-offspring. The differences which did occur showed fathers to be more projecting and identified but mothers to have more actual awareness of their sons' beliefs. These results were inconsistent with the concept of the schizophrenogenic parent, especially the father, as presented in the literature and suggested a need for reformulation.

2. The parents of chronically ill in contrast with acutely ill patients appeared generally to maintain greater psychological distance from their sons, being less identified and aware but more projecting in their personal beliefs. This finding was interpreted in terms of a reciprocal effect of the intensity of the disorder and the habitual modes of response which parents have in relating to their sons.

REFERENCES

1. ABERELE, D. F., & NEAGELE, K. D. Middle class fathers' occupations, role and attitude toward children. *Amer. J. Orthopsychiat.*, 1952, **22**, 366-378.
2. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950.
3. ARIETI, S. Interpretation of Schizophrenia. New York: Robert Brunner, 1955.
4. BASSEL, J. S. A study of the attitudes of schizophrenic patients toward mental illness: an investigation of the differences in certain stereotyped attitudes regarding mental illness in schizophrenic patients referred for release from the hospital, and in those who had to remain in the hospital. Unpublished doctor's dissertation, New York Univer., 1955.
5. CASS, L. K. An investigation of some important variables in the parent-child relationship. Unpublished doctor's dissertation, Ohio State Univer., 1950.
6. ———. Parent-child relationship and delinquency. *J. Abnorm. & Soc. Psychol.*, 1952, **47**, 101-104.
7. DESPERT, L. Schizophrenia in children. *Psychiat. Quart.*, 1938, **12**, 336-371.
8. DIETHELM, O. Treatment in Psychiatry. Springfield, Ill.: Charles C Thomas, 1950.
9. DWORIN, J., & WYANT, O. Authoritarian patterns in the mothers of schizophrenics. *J. Clin. Psychol.*, 1957, **13**, 332-338.
10. GERARD, D. L., & SIEGEL, J. The family background of schizoprhenia. *Psychiat. Quart.*, 1950, **24**, 47-73.
11. HAJDU-GIMES, L. Contributions to the etiology of schizophrenia. *Psychoanal. Rev.*, 1940, **27**, 421-438.
12. KRINSKY, L. Parent-patient relationships as reflected in personal beliefs: A comparison of fathers and mothers of acutely and chronically ill schizophrenics. Unpublished doctor's dissertation, Adelphi College, 1958.

13. LANE, R. C. Familial attitudes of paranoid schizophrenics and normal individuals of different socio-economic levels. Unpublished doctor's dissertation, New York Univer., 1954.

14. LEVINSON, S., & WHITNEY, M. Z. Sessions with relatives of mental hospital patients. *Ment. Hyg., N.Y.*, 1955, **39**, 118-125.

15. MARK, J. C. The attitude of the mothers of male schizophrenics toward child behavior. *J. Abnorm. & Soc. Psychol.*, 1953, **48**, 135-139.

16. NIELSEN, J. M., & THOMPSON, G. W. The Engrammes of Psychiatry. Springfield, Ill.: Charles C Thomas, 1947.

17. PROUT, C. T., & WHITE, MARY A. A controlled study of personality relationships in mothers of male schizophrenic patients. *Amer. J. Psychiat.*, 1950, **107**, 251-256.

18. REICHARD, S., & TILLMAN, C. Pattern of parent-child relationships in schizophrenia. *Psychiatry*, 1950, **13**, 247-257.

19. SINGER, J. L. Projected familial attitudes as a function of socio-economic status. *J. Consult. Psychol.*, 1954, **18**, 99-104.

20. TERRY, J., & RENNIE, T. Analysis of paregasia. *Nerv. Ment. Dis. Monogr.*, 1938, **64**.

21. TIETZE, T. A study of mothers of schizophrenic patients. *Psychiatry*, **12**, 1949, 55-65.

22. WYLIE, P. Generation of Vipers. New York: Farrar and Rhinehart, 1942.

South Oaks Psychiatric Hospital
Sunrise Highway
Amityville, Long Island, New York

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 399-407.

PERCEPTUAL THRESHOLD DECREMENTS AS A FUNCTION OF SKILL AND CHANCE EXPECTANCIES*[1]

Department of Psychology, Kansas State University

E. JERRY PHARES

A. INTRODUCTION

Recently, considerable attention has been devoted in psychology to an analysis of the situation as it affects human behavior. Masling (11) has reviewed the influence of situational factors on projective testing. Likewise, Rotter (18) has applied a social learning theory[2] analysis to test procedures. More general emphasis on the situation has been urged by Brunswik (3), Helson (5), and Rotter (17).

The present study will focus on one particular situational variable. That is, the extent to which situations are categorized by *S*s as ones where obtained reinforcements are a function either of their skill and ability or a function of chance, non-controllable factors. Evidence is accumulating that *S*s will behave differently in situations where reinforcement is contingent upon some ability or characteristic of self as opposed to situations where chance, non-predictable, or uncontrollable forces operate.

Phares (12) studied the skill-chance distinction by examining the effect of such situations on expectancy changes. The major findings were that a chance situation produced smaller and less frequent expectancy changes than a skill situation when the tasks involved color and line matching activities. James (7) essentially replicated Phares' results and also demonstrated that a chance situation leads to faster extinction of expectancies, less generalization of expectancies from one task to another, and greater recovery following extinction. James and Rotter (8) studied the differential effects of partial versus 100 per cent reinforcement schedules under skill and chance conditions and concluded that resistance to extinction of expectancies was greater with 100 per cent reinforcement than with 50 per cent reinforcement under skill conditions. However, under chance conditions, partial reinforcement was more

---

* Received in the Editorial Office on January 12, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was supported by the National Institute of Mental Health, U.S. Public Health Service, under Research Grant M-3852.

[2] Social learning theory (Rotter, 1954) provided the general framework within which this study was conducted.

effective in retarding the effects of extinction of expectancies. Rather than relying on experimental instructions, Rotter, Liverant, and Crowne (19) have recently produced results similar to those of the foregoing three studies by using tasks differing in the extent to which they aroused skill or chance expectancies. Feather (4) found that attainment attractiveness of goals are inversely related to success probability in both ego-related and chance situations but the relationship is less apparent in the chance situation. In studying probability preferences, Littig (9) could demonstrate no significant differential preferences in chance versus skill situations but concluded that it would be unwise, on the basis of his results, to neglect the skill-chance dichotomy.

The measurement of individual differences in the generalized tendency to categorize events as being determined by skill or internal factors as opposed to chance or external factors has also been explored. Some evidence exists for such differences (7, 10, 12, 20) but the results must be considered highly tentative as yet (22).

Thus far, research on the skill-chance situational variable has dealt primarily with expectancies, goal values, and probability preferences. The present study attempts to relate this situational variable to perceptual behavior. Specifically, we shall attempt to show that when escape from a painful stimulus (shock) is possible only on a chance basis, the difference between pre- and post-experiment recognition thresholds for shock-associated syllables will be smaller than in a skill situation where escape definitely depends on *S*'s ability. Thus, when *S* develops an expectancy that his skill in a recognition-learning situation will lead to the positive goal of escape, it is to his advantage to attend, thus enabling lower visual thresholds. On the other hand, where escape is possible only on a chance or random basis, no such expectancy will develop thus leading to a significantly smaller decrease in thresholds. This hypothesis seems consistent with the analysis of similar situations made by Solley and Murphy (21). We are essentially predicting "vigilance" as a function of skill cues. Actual "repression" in the chance situation will not likely result since escape can occur occasionally as a matter of luck.

Reece (14) demonstrated that escapability from shock leads to relatively lower visual thresholds for shock-associated stimuli than does non-escapability. However, the factor of predictability of shock (shock paired consistently with the same stimuli) did not relate to differences in recognition thresholds due, perhaps, to a failure to establish the predictability of shock effectively. Similar results have been reported by Rosen (15) who predicted that when perception of a stimulus is painful to *S* and direct avoidance of the stimulus is impossible, the recognition threshold for it will rise. When perception of the stimulus aids *S* in his avoidance, the threshold for it will become lower. Pustell (13) showed sex differences in tendency toward vigilance or defense. That is, males attempting to escape from punishment showed vigilance while females passively submitted and showed defense. Reviews of the literature in this general area may be found in Allport (1) and Inglis (6).

Thus far, the evidence points to the fact that *S*s will react with relatively low perceptual thresholds to situations where escape from noxious stimulation is possible through skilled responses on their part. Furthermore, where no such escape is possible, relatively high thresholds ensue or even "repression." Chance situations, however, allow for escape occasionally but not as a function of anything *S* does. Chance situations would seem to represent a point on the avoidable-non-avoidable continuum somewhere between the two extremes. Several studies have successfully compared perceptual thresholds in the two extreme situations. The goal of the present study is to specifically relate the skill-chance variable to changes in visual thresholds for pain-associated stimuli; a variable which is not the same as simple potential for avoidance or non-avoidance of punishment.

B. METHOD

1. *Subjects*

*S*s were 21 male and 21 female paid volunteers. The skill, chance, and non-shock groups were each composed of seven males and seven females. Most of the *S*s were recruited from undergraduate classes at Kansas State University, although a few were non-campus or graduate student volunteers.

2. *Procedure and Apparatus*

Each *S* was initially seated before a Gerbrands mirror tachistoscope. He was shown successively a total of 12 CVC trigrams (2) in order to establish pre-experimental thresholds. Syllables were printed on cards in 8 point century caps. Six of the syllables were later associated with electric shock, while the other six were never so associated. The two groups of syllables were matched in association value.

Next, *S* was brought into the experimental room and seated before a reaction time panel which contained six buttons. After he was comfortable the remainder of the apparatus was briefly described to him. Next, it was explained that it was necessary to establish the level of shock which he found painful. The electrodes were then strapped to the palmar surface of the first finger of his non-dominant hand. The shock was produced by an Applegate 226B stimulator. *S* then rated a series of shocks on a seven point scale from "unable to feel" to "painful." That level of shock which he rated as painful was used throughout the experiment for him, except that initially the level used was slightly below that of the painful rating and the level was then gradually increased over trials to counteract adaptation.

Following this procedure both skill and chance *S*s were given the following instructions:

> You will be given 10 trials. On each trial, 12 nonsense syllables will be projected on the wall over there, one at a time. The same syllables will be shown each trial, but the order in which they appear will be random from trial to trial. Six of these syllables are shock syllables—that is, every time one of these six appear on the wall, an electric shock will simultaneously occur. The shock will stay for the entire duration of the syllable's exposure, which is two seconds. The other six syllables will never be associated with shock. In other words, six of the 12 syllables will always be accompanied by shock—the other six never accompanied by shock, regardless of the order from trial to trial. Okay?

For skill *S*s the following additional instructions were given:

> Now here is the important part. Each of the six shock syllables has its own control button on your panel there. This never changes throughout the experiment. For example, it might be that the first shock syllable goes with button number 4. If so, it will always go with that button. Herein lies the importance of this. If you press the correct button for a given shock syllable, the shock will terminate at that point. In other words, by learning which shock syllable goes with which button, you can almost completely escape the shock associated with that syllable. In short, the extent of shock you receive is dependent on your skill in figuring out which button goes with which shock syllable. Potentially, then, you are in control of the situation. Okay?

For chance *S*s the following material was substituted:

> Now here is the important part. Whenever a shock syllable comes on you may, by pressing the correct button, terminate the shock at that point. However, which button is the correct one for each syllable from time to time is determined by me. As far as you are concerned it is a guessing game. For example, the first time VEQ appears, button number 4 may be correct. But the next time VEQ appears, button number 2 may be the correct one. When you guess correctly, the shock terminates but when your guess is incorrect, the shock stays on for the entire duration of that syllable's exposure. Thus, it is all a matter of chance or luck as far as you are concerned. Okay?

All *S*s were also instructed to read each syllable aloud to ensure that they actually attended to the syllables thus preventing chance *S*s from concentrating solely on button pressing.

E was seated before a control panel which was connected with *S*'s panel and could thus tell which button was pressed each time. If the correct button was pressed *E* flipped a manual switch terminating the shock. Another *E* changed slides for the projector.

All 14 skill *S*s were run first. Then each chance *S* was successively matched with a skill *S* so that both sex and the number of escapes and point of escape in the trial sequence were identical for both groups. Thus, the only real difference between skill and chance groups was the instructions since both groups escaped an equal number of times and on the same syllables at the same point in the sequence. Whatever differences might be obtained in their thresholds would thus have to be a function of differential expectancies for control of the situation.

A total of 10 trials were run with each trial consisting of the exposure of the 12 syllables in a random order. The order was, however, constant for all *S*s.

Following this procedure, all *S*s were returned to the tachistoscope and their recognition thresholds again taken for the 12 syllables. For the non-shock group, both pre- and post-experimental thresholds were obtained. In between, however, they merely viewed the 120 slides without the accompanying shock.

C. Results and Discussion

It was predicted that the decrement in recognition thresholds for shock-associated stimuli would be greater under conditions of skill than under chance conditions. In order to control for individual differences in initial thresholds, ratios of the differences between pre- and post-experimental thresholds to initial thresholds were used for each *S*. Reference to Tables 1 and 2 indicates that for shock syllables the difference in decrements be-

TABLE 1
Before and After Perceptual Scores and Their Differences for Shock and Non-Shock Syllables

Group	N	Before Mean	Before SD	After Mean	After SD	Diff.
			Shock Syllables			
Skill	14	30.4*	12.9	20.4	13.4	10.0
Chance	14	34.9	9.7	29.3	12.0	5.6
Non-shock	14	35.2	8.9	24.2	9.1	11.0
			Non-shock Syllables			
Skill	14	28.5	14.5	21.4	11.2	7.1
Chance	14	33.2	11.4	30.5	14.3	2.7
Non-shock	14	33.0	9.5	24.8	10.0	8.2

* All scores expressed in 100ths of a second.

tween the skill and chance groups is .17 and significant in the predicted direction. The same difference between skill and non-shock groups is .02 and not significant. For non-shock versus chance groups the difference is .15, significant again in the predicted direction. Therefore, as expected, skill *S*s showed a greater decrement in thresholds than did chance *S*s. Indeed, chance *S*s were inferior to control *S*s. It appears, then, that in situations where *S* is rewarded by escape from a noxious stimulus on the basis of his skill in learning, a sort of vigilance develops as compared to situations where escape follows a behavior only on a chance basis.

TABLE 2
SIGNIFICANCE TESTS OF RATIO SCORES BETWEEN SKILL, CHANCE, AND NON-SHOCK GROUPS FOR SHOCK AND NON-SHOCK SYLLABLES

		Shock Syllables					Non-Shock Syllables				
Group	N	Mean	Diff.	SE_{diff}	t	p	Mean	Diff.	SE_{diff}	t	p
Skill	14	.33*					.25				
			.17	.075	2.27	.01**		.17	.067	2.54	.008
Chance	14	.16					.08				
			.15	.075	2.00	.025		.17	.067	2.54	.008
Non-Shock	14	.31					.25				

* Mean scores refer to ratios of the difference between pre- and post-experimental thresholds to initial thresholds.

** Based on one-tailed tests.

Comparisons between Skill and Non-Shock Groups were non-significant.

The fact that there were no significant differences between skill and non-shock groups replicates a previous finding by Reece (1954). It is probable that even though the skill group had a source of reinforcement or incentive through escape, the shock also produced disruptive effects. These two effects, as Reece argues, would likely cancel each other. The non-shock group, on the other hand, was subject neither to reinforcement nor disruption. Therefore, the net effect is two groups performing similarly. The chance group is subject to disruption but expects very little reinforcement except as a function of luck.

Analyses similar to the above were also conducted for non-shock syllables. As before, the skill group showed a significantly greater decrement than the chance group (diff. = .17). There was no difference between skill and non-shock groups, but again, the difference of .17 between non-shock and chance groups was significant. These results suggest that the same processes operated here as in the case of the shock syllables to produce similar results. Also, *t* tests yielded no significant differences between the decrements for shock syllables and neutral syllables within any of the three conditions.

This would indicate, as in Reece's study (14), stimulus generalization prevented any differences between shock and non-shock syllables from appearing.

The foregoing results strongly suggest that behavior-reward sequences experienced under skill conditions result in perceptual behavior different from that developed under chance conditions. *S*s who feel they have control of the situation are likely to exhibit perceptual behavior that will better enable them to cope with potentially threatening situations than *S*s who feel chance or other non-controllable forces determine whether their behavior will be successful. Thus, there would appear to be a direct relationship between extent of coping behavior and degree of expectancey that one's skill or ability is the crucial variable in achieving reinforcement. Furthermore, the fact that the above results were true also of non-shock stimuli suggests the generalization factor. That is, expectancies developed in one situation or for one class of stimuli may generalize to other related situations or stimuli thus rendering the whole concept of the skill-chance dichotomy more important.

These results appear quite consistent, generally, with other studies dealing with the variables of escapability and avoidability of noxious stimulation (14, 15). However, chance situations do not necessarily imply inability to escape or avoid, but rather, an inability to control by self-initiated behaviors the occurrence or non-occurrence of reinforcements. Perhaps situations may be ordered on a continuum with respect to the extent to which noxious stimulation may be avoided. Avoidable or skill situations would be at one extreme and would lead to a kind of perceptual vigilance. Situations in which pain is unavoidable despite any behavior on the part of *S* would constitute the other extreme and lead to perceptual defense or repression of the painful stimulus. At a point somewhere in between would be situations where *S*s' behavior would lead to escape only on a chance basis, perhaps leading to perceptual behavior significantly less vigilant than in the skill situation but also less defensive than in the completely unavoidable situation.

D. SUMMARY

This experiment was designed to test the hypothesis that when escape from a painful stimulus as a result of *S*'s behavior is possible only on a chance basis, the difference between pre- and post-experiment recognition thresholds for shock-associated stimuli will be smaller than in a skill situation where escape depends on *S*'s ability. This prediction was based on the rationale that an expectancy for control of the situation by *S* will lead him to behave in a manner most likely to capitalize on his ability to so control

the situation and thus minimize pain. Since, in the chance situation, *S* is not in control of the situation, significantly smaller decreases in thresholds will occur.

Three groups of *S*s, 14 skill, 14 chance, and 14 non-shock *S*s were used. Tachistoscopic thresholds for 12 nonsense syllables were taken for all *S*s. Following this, the skill group was given ten trials in which six of the syllables were always accompanied by shock and the other six were never accompanied by shock. *S* could, however, by learning which shock syllable was associated with which button on his control panel, terminate the shock each time. Chance *S*s were told that the correct button changed continuously so that no learning was possible and thus escape occurred only on a random basis. Chance *S*s were matched with skill *S*s with respect to number of escapes, syllables on which escape occurred, and sex. Following this procedure, thresholds were again taken for each of the 12 syllables. For non-shock *S*s, the pre- and post-experiment thresholds were taken while in between, the series of syllables were exposed for ten trials without shock.

The results indicated that, as predicted, threshold decrements were significantly greater for skill *S*s than for chance *S*s for both shock and non-shock syllables. There were no significant differences between skill and non-shock groups although the latter group showed smaller decrements for the non-shock syllables. The non-shock group exhibited greater decrements than did the chance group for both classes of syllables. The difference in decrements between shock and non-shock syllables was not significantly different for any of the three conditions thus indicating the operation of generalization factors.

Relationships between these results and perceptual vigilance and defense studies were discussed along with other implications of the skill-chance situational dichotomy.

REFERENCES

1. ALLPORT, F. H. Theories of Perception and the Concept of Structure. New York: Wiley, 1955.
2. ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.,* 1960, **74** (whole No. 497).
3. BRUNSWIK, E. The probability point of view. *In* M. H. Marx, (*Ed.*), *Psychological Theory.* New York: Macmillan, 1951.
4. FEATHER, N. T. Success probability and choice behavior. *J. Exp. Psychol.,* 1959, **58**, 257-266.
5. HELSON, H. Adaptation level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.,* 1948, **55**, 297-313.
6. INGLIS, J. Abnormalities of motivation and "ego-functions." *In* H. J. Eysenck (*Ed.*), *Handbook of Abnormal Psychology.* New York: Basic Books, 1961.

7. James, W. H. Internal versus external control of reinforcement as a basic variable in learning theory. Unpublished doctoral dissertation, Ohio State Univer., 1957.

8. James, W. H., & Rotter, J. B. Partial and 100 per cent reinforcement under chance and skill conditions. *J. Exp. Psychol.,* 1958, **55**, 397-403.

9. Littig, L. W. The effect of skill and chance orientations on probability preferences: A preliminary report. Paper read at Midwestern Psychological Assoc., Chicago, May, 1961.

10. Liverant, S., & Scodel, A. Internal and external control as determinants of decision making under conditions of risk. *Psychol. Rep.,* 1960, **7**, 59-67.

11. Masling, J. The influence of situational and interpersonal variables in projective testing. *Psychol. Bull.,* 1960, **57**, 65-85.

12. Phares, E. J. Expectancy changes in skill and chance situations. *J. Abnorm. & Soc. Psychol.,* 1957, **54**, 339-342.

13. Pustell, T. E. The experimental induction of perceptual vigilance and defense. *J. Pers.,* 1957, **25**, 425-438.

14. Reece, M. M. The effect of shock on recognition thresholds. *J. Abnorm. & Soc. Psychol.,* 1954, **49**, 165-172.

15. Rosen, A. C. Change in perceptual threshold as a protective function of the organism. *J. Pers.,* 1954, **23**, 182-194.

16. Rotter, J. B. Social Learning and Clinical Psychology. New York: Prentice-Hall, 1954.

17. ———. The role of the psychological situation in determining the direction of human behavior. *In* M. R. Jones (*Ed.*), *Nebraska Symposium on Motivation,* 1955. Lincoln: Univer. Nebraska Press, 1955.

18. ———. Some implications of a social learning theory for the prediction of goal directed behavior from testing procedures. *Psychol. Rev.,* 1960, **67**, 301-316.

19. Rotter, J. B., Liverant, S., & Crowne, D. P. The growth and extinction of expectancies in chance controlled and skilled tasks. *J. Psychol.,* 1961, **52**, 161-177.

20. Rotter, J. B., Seeman, M., & Liverant, S. Internal versus external control of reinforcements: A major variable in behavior theory. *In* N. Washburne (*Ed.*), *Decisions, Values, and Groups,* Vol. II. London: Pergamon Press, 1961.

21. Solley, C. M., & Murphy, G. Development of the Perceptual World. New York: Basic Books, 1960.

22. Straley, G. L., & Phares, E. J. Tachistoscopic and level-of-aspiration measures of the internal-external control variable. Unpublished manuscript, Kansas State Univer., 1961.

Department of Psychology
Ohio State University
1945 North High Street
Columbus 10, Ohio

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 409-415.

SOME PERSONALITY CORRELATES OF VERBAL CONDITIONING*

W. ROBERT HETRICK AND KURT HAAS[1]

Psychology Department, Bowling Green State University

A. INTRODUCTION

While relatively few studies have probed the relationship between personality attributes and verbal conditioning, the investigations that have been conducted have frequently produced conflicting results. It appears, however, as if those personality variables that could be logically related to verbal conditioning have most frequently been confirmed as exercising significant influence on the conditioning process (2, 16, 17, 21). Considerable evidence has been adduced also, which suggests some underlying structural similarity between psychotherapy and verbal conditioning (23, 24). For example, it has been shown that anxiety is both related to success in therapy (8, 9, 10, 11, 18, 20), and to conditionability (25). Aggression on the other hand, seems to be detrimental to both processes (1, 7, 15).

The purpose of this study is to investigate the relationships of select personality variables on performance in a verbal conditioning task. The personality variables chosen for study were done so because they empirically correlated with psychotherapy and hence, assumedly, were logically related to verbal conditioning.

Barron (4), and Wirt (27) found ego-strength to be positively related to improvement in psychotherapy. Ego-strength, as defined by Barron, can be conceived as having a facilitating influence on the socially interactive process of verbal conditioning.

Depression, another personality variable studied, has been found to be negatively related to improvement and continuation in psychotherapy (5, 8). The characteristics manifested by a more depressed person may be viewed as being detrimental to a socially interactive process such as verbal conditioning.

From the descriptions of the psychopathic personality (6), and the work of some investigators, it could be expected that the personality traits of the

---

* Received in the Editorial Office on January 16, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was supported, in part, by funds from the Committee for Scholarly Advancement, Bowling Green State University.

psychopath would be detrimental to performance in verbal conditioning, as well as in therapy (12, 14).

This study then tested the predictions that ego-strength scores will show a positive relationship with verbal conditioning and that depression and psychopathy will evidence a negative relationship.

B. METHOD

1. *Apparatus and Subjects*

A questionnaire made up of the questions constituting the D, Pd, and Es scales of the MMPI was devised. In addition, 120 stimulus cards, 3 × 5, on which a verb in simple past tense was typed in the card's center, and below which the pronouns, I, We, You, He, She, and They were typed in a line, was used to elicit conditioning. The pronouns appeared in 120 different orders while each pronoun appeared in each position with approximately equal frequency.

The *S*s used in this study were 78 college student volunteers from elementary psychology classes at a State University.

2. *Procedure*

The instructor of each psychology class used in this study administered the personality inventory to all *S*s. The students who desired to participate in the second part of the study [conditioning], which purportedly investigated the "flow of language," were asked to volunteer.

Each conditioning session was begun with instructions to the *S* to quickly form a sentence using the verb and a pronoun of choice, presented on the card. For Trials 31 through 120, the acquisition trials, the experimenter responded with the word "good" in a flat, unemotional tone at the end of each sentence using the pronouns which were predetermined to be reinforced. Half of the *S*s received reinforcement to the pronouns I and We while the other half received reinforcement to the pronouns He and They. This was done in order to partially control for possible interactions between the class of pronouns reinforced (self-referent and other-referent) and the personality variables.

At the end of the conditioning session, each *S* was informally questioned in order to determine if he had a set to respond in any biasing manner or if he were aware of the reinforcing contingency. The protocols of *S*s who were determined to have had a pronoun or position preference were excluded.

C. RESULTS

Pearson product moment correlation coefficients were computed between the frequency of the reinforced and nonreinforced responses for Trials 1 through 30 (operant trials) and the *S*s' verbal conditioning scores. These correlation coefficients were insignificant which indicates that the operant level of the *S* did not significantly influence his verbal conditioning performance regardless of the pronoun class compared.

The means and standard deviations of the 78 subjects on the Es, D, and Pd scales are presented separately in Table 1 for the I-We and He-They reinforced groups and for normative samples. It is apparent that the scores earned by the *S*s used in this study are within the expected limits as compared with the norms reported in the literature (13).

TABLE 1
MMPI SCORES FOR SUBJECTS IN THE I-WE AND HE-THEY REINFORCEED GROUPS AND NORMATIVE DATA

Scale (normative sample)	N	Group	M	SD
Ego-Strength (ES)	39	I-We	44.82	6.33
	39	He-They	44.82	3.96
	78	Total	44.82	5.28
(Hathaway & Briggs, 1957)	226	MMPI Norm sample	44.33	6.21
(Barron, 1953a)	33	Standardization sample	41.94	13.30
Depression (D)	39	I-We	20.59	5.51
	39	He-They	20.39	4.54
	78	Total	20.49	5.05
(Hathaway & McKinley, 1956)	690	Normals	18.14	4.90
(Hathaway & (McKinley, 1956)	33	Depressed group	36.68	4.70
Psychopathy (Pd)	39	I-We	17.95	5.16
	39	He-They	16.03	4.54
	78	Total	16.99	4.95
(McKinley & Hathaway, 1956)	397	Normal Females	13.44	4.23
(McKinley & Hathaway, 1956)	78	Clinic cases	21.44	6.23

Table 2 shows the frequency of responses which were either reinforced or not reinforced. When the *S*s were not reinforced for using the pronouns

I and We, or He and They, there was no significant increase in the frequency of the pronouns' occurrence from the first to the last 30 trial groups. However, when reinforcement was given to these pronouns there was a significant increase in the frequency of their occurrence. This data indicates that verbal conditioning occurred.

TABLE 2
FREQUENCIES OF I-WE AND HE-THEY RESPONSES FOR REINFORCED AND UNREINFORCED GROUPS

Group		Trials 1-30	91-120	*t*
		Reinforced		
I-We	M	11.77	14.13	2.10*
	SD	3.97	7.65	
He-They	M	10.41	13.67	4.14**
	SD	2.99	4.21	
		Unreinforced		
I-We	M	11.74	9.36	—
	SD	3.29	4.29	
He-They	M	10.59	10.92	.38
	SD	3.24	5.86	

* Significant at the 5% level.
** Significant at the 1% level.

Differences between the verbal conditioning scores for the two reinforced groups (I-We, He-They) were not significant ($t = .15$). Thus, the class of pronouns reinforced had no significant differential influence on the verbal conditioning scores.

The Pearson product moment correlations between verbal conditioning scores and the three personality variables (ego-strength, $r = .08$; depression, $r = .01$; psychopathy, $r = .07$) were very low and did not approach significance. Similarly, scatter plots for the same scores clearly indicated that the relationships between the measures were linear but negligible.

D. DISCUSSION

The correlations between *S*s' scores on the Es, D, and Pd, scales of the MMPI and the verbal conditioning task were low and failed to reach statistical significance. Perhaps the personality variables investigated do not meaningfully influence verbal conditioning performance. On the other hand, certain aspects of the experimental design used to examine the hypotheses may have obscured any possible relationships that might exist. For instance, some personality variables seem to be influential at different levels of intensity (19), or with different kinds of *S*s (22, 25). It may be that when an *S* functions at the extremes of a personality dimension, he performs differentially in a verbal conditioning task. The *S*s used in this study were essentially within normal limits on each personality variable investigated so that the trait may have had a negligible effect on their conditioning performance.

Babladelis (2) emphasized another important consideration by pointing out that the verbal conditioning process may be very different from the psychotherapeutic process. In this study, for example, the verbal conditioning paradigm was a structured, non-interview one, while the psychotherapeutic process is an interview situation which is by comparison much less structured. Because personality traits probably operate differently in such dissimilar situations, the relationships they have to each situation may vary markedly. For the purposes of this study a verbal conditioning process more similar to the psychotherapeutic process (conditioning in an interview situation) such as those used by some investigators (26) might have been more appropriate.

E. SUMMARY

Reviews of the verbal conditioning literature indicated that relatively few investigations of the personality correlates of verbal conditioning have been undertaken. This investigation sought to test the predictions, suggested by previous studies, that because of their relationship to psychotherapy, scores on an ego-strength measure would correlate positively with scores in a verbal conditioning task, while scores on measures of depression and psychopathic deviation would correlate negatively with verbal conditioning.

Seventy-eight students from undergraduate courses in psychology filled out a personality inventory (MMPI) which consisted of the Es, D, and Pd scales. Within eight days after the administration of the personality inventory, each *S* was individually conditioned. As the *S*s made up sentences using verbs and pronouns presented on index cards, one-half of them were reinforced when sentences were begun with the pronouns I and We, and the remaining half when sentences were begun with the pronouns He and They.

Analyses of the data indicated that the *S*s in each reinforcement group were verbally conditioned but that none of the three personality measures correlated significantly with the verbal conditioning performance. Some aspects of the experimental design and apparent dissimilarities between psychotherapy and the present conditioning situation were suggested as probable explanations for the failure of the results to confirm the predictions.

REFERENCES

1. Albee, G. W. Patterns of aggression in psychopathology. *J. Consult. Psychol.*, 1950, **14**, 465-468.
2. Babladelis, G. Personality and verbal conditioning effects. *J. Abnorm. & Soc. Psychol.*, 1961, **62**, 41-43.
3. Bandura, A. Psychotherapy as a learning process. *Psychol. Bull.*, 1961, **58**, 143-159.
4. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. Consult. Psychol.*, 1953, **17**, 327-333.
5. ———. Some test correlates of response to psychotherapy. *J. Consult. Psychol.*, 1953, **17**, 235-241.
6. Cleckley, H. The Mask of Sanity. St. Louis: C. V. Mosby Co., 1955.
7. Feldman, D. A., Pascal, G. R., & Swenson, C. H. Direction of aggression as a prognostic variable in mental illness. *J. Consult. Psychol.*, 1954, **18**, 167-170.
8. Frank, J. E., Gleidman, L. H., Imber, S. D., Nash, E. H., & Stone, A. R. Why patients leave psychotherapy. *A.M.A. Arch. Neurol. Psychiat.*, 1957, **77**, 283-299.
9. Gallagher, J. J. Manifest anxiety changes concomitant with client-centered therapy. *J. Consult. Psychol.*, 1953, **17**, 443-446 (a).
10. ———. Test indicators for therapy prognosis. *J. Consult. Psychol.*, 1954, **18**, 409-413.
11. Gibby, R. G., Stotsky, B. A., Hiler, E. W., & Miller, D. R. Validation of Rorschach criteria for predicting duration of therapy. *J. Consult. Psychol.*, 1954, **18**, 185-191.
12. Harris, R E., & Christiansen, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269-284.
13. Hathaway, S. R., & Briggs, P. F. Some normative data on MMPI scales. *J. Clin. Psychol.*, 1957, **13**, 364-368.
14. Hunt, J. McV., Ewing, T. N., LaForge, R., & Gilbert, W. M. An integrated approach to research on therapeutic counseling with samples of results, *J. Counsel. Psychol.*, 1959, **6**, 46-54.
15. Kirman, W. J. The relationship of learning, with and without awareness, to personality needs. *Dissertation Abstr.*, 1958, **19**, 362-363.
16. Krasner, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148-170.
17. Krasner, L., Ullmann, L., Weiss, R., & Collins, B. A personality correlate of individual differences in responsivity to verbal conditioning as obtained by three examiners. *J. Clin. Psychol.* (in press).
18. Lorr, M., Katz, M., & Eli, A. The prediction of length of stay in psychotherapy. *J. Consult. Psychol.*, 1958, **22**, 321-327.
19. Rogers, J. Operant conditioning in a quasi-therapy setting. *J. Abnorm. & Soc. Psych.*, 1960, **60**, 247-252.
20. Rosenberg, S. The relationship of certain personality factors to prognosis in psychotherapy. *J. Clin. Psychol.*, 1954, **10**, 341-345.
21. Salzinger, K. Experimental manipulation of verbal behavior: A review. *J. Gen. Psychol.*, 1959, **61**, 65-94.
22. Sarason, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. Abnorm. & Soc. Psychol.*, 1958, **56**, 339-344.
23. Shaw, F. S. Some postulates concerning psychotherapy. *J. Consult. Psychol.*, 1948, **12**, 426-432.

24. SHOBEN, E. J., JR. Psychotherapy as a problem in learning theory. *Psychol. Bull,* 1959, **46**, 355-392.

25. TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. Abnorm. & Soc. Psychol.,* 1955, **51**, 496-501.

26. ULLMANN, L., KRASNER, L., & COLLINS, B. Modification of behavior through verbal conditioning: effects in group therapy. *J. Abnorm. & Soc. Psychol.,* (in press).

27. WIRT, R. D. Further validation of the ego-strength scale. *J. Consult. Psychol.,* 1955, **19**, 444.

Larned State Hospital
Larned, Kansas

Psychology Department
Bowling Green State University
Bowling Green, Ohio

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 417-424.

STIMULUS GENERALIZATION WITH AGGRESSIVE VERBAL STIMULI: A NEW PARADIGM*[1]

Department of Psychology, University of Pittsburgh

ARNOLD H. BUSS

A. INTRODUCTION

The purpose of this study was two-fold. First, it attempted to test the generality of previous findings concerning stimulus generalization with aggressive verbal stimuli. Second, it tested a new method for studying stimulus generalization.

A previous study (1) demonstrated that there is a tendency to match the intensity of a vocal response to the intensity of aggressive words. Direction of generalization was also found to be important. When generalizing from intense to mild aggressive stimuli, shouting the words steepens the generalization gradient and whispering flattens it. When generalizing from mild to intense aggressive stimuli, shouting flattens the gradient and whispering steepens it. These findings hold whether the response is vocal or motor (3). Will consistent results be found when the stimulus generalization paradigm is altered?

B. METHOD

1. *Stimuli*

The stimuli are words that have been scaled for intensity of aggression (2). They were divided into five lists of 20 each on the basis of intensity scale values. The range of scale values for the word groups was as follows:

1	1.8 - 3.5
2	4.0 - 4.5
3	5.0 - 5.5
4	5.7 - 6.7
5	7.0 - 8.8

Note that there is no overlap between lists. It was necessary to vary the range within lists in order to prevent overlap between lists. There are more words in the middle of the scale than at the ends, which means that the range of the middle groups must be smaller.

---

* Received in the Editorial Office on January 29, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported by Research Grant M-2861 from the National Institute of Mental Health, United States Public Health Service.

2. *Procedure*

The stimuli were presented five at a time in the learning series. The five words on any given trial consisted of one member from each of the five lists of words:

1	2	3	4	5
competitive	pest	anger	loathing	war

Since a learning series of 100 trials was necessary, words from each of the five lists were recombined, each word being used a total of five times. A given word never reappeared with its previous partners. Thus the next time *competitive* appeared, the other words on the trial were:

bruise	browbeating	cruel	annihilation

The words were presented in random order on a given trial, and the sequence of presentation from trial to trial was randomized, with the limitation that no word be repeated until all the other members of its intensity group had been presented.

The words were presented by fives (one from each intensity list) on a memory drum. *S* was given the following instructions.

> I am going to show you a series of words. They will be presented five at a time, and each group of five words will appear for eight seconds. What you have to do is read *one* of the words aloud. Now you must make up your mind within eight seconds which word to read and then say it aloud. I will tell you whether you are right or wrong, and in that way you will get the hang of it.

Half the *S*s were reinforced for saying the most intense word; if they said one of the four milder alternatives, they were told *Wrong*. The other half were reinforced for saying the mildest word; if they said one of the four more intense words, they were told *Wrong*.

The criterion of learning was five consecutive correct responses. If *S* did not reach criterion by 100 trials, he was dropped from the experiment.

At the end of the learning series *S* was given the following instructions.

> Now we are going to switch over here (second memory drum), but I want you to keep going the way you have been. The only difference is that now there are four words in each group. I will no longer tell you whether you are right or wrong.

Then the generalization series commenced. This series differed from the learning series in two ways. First, there were 80 trials rather than 100. Second, there were only four words per trial. The words that had been reinforced in the learning series did not appear in the generalization series. *S*s who were reinforced for List 5 (most intense words) did not see these words in the generalization series. Similarly, *S*s who were reinforced for List 1 (mildest words) did not see these words in the generalization series.

In the absence of the words that he had been reinforced for saying, *S* was forced to choose one of the four remaining (previously unreinforced) words. He might simply transpose, saying only the next most intense word (List 4) or the next mildest word, depending upon whether List 1 or List 5 had been reinforced during learning. If this were to occur, there would be no generalization, merely transposition; the paradigm would be useless for investigating stimulus generalization. The success of the paradigm requires that there be a spread of response from the reinforced group of words to the other four groups.

3. *Design*

The first independent variable, direction of generalization, has already been mentioned. Half the *S*s generalized from intense to mild stimuli, and half generalized from mild to intense stimuli.

The second independent variable was intensity of the reinforced vocal response. One group of *S*s was instructed to shout the words and was given preliminary practice to insure shouting. One group was instructed to say the words aloud in a normal speaking voice. A third group was instructed to whisper the words. Thus the three groups were Shout, Aloud, and Whisper.

The last independent variable was sex of *S*. Half the *S*s were men, and half were women.

4. *Subjects*

*S*s were college students recruited from Introductory Psychology classes. A total of 235 *S*s took part in the experiment. Of these, 180 reached criterion, and 55 did not. There were no differences between the experimental groups or the sexes in the number of non-learners. However, direction of generalization was important, for it was easier to learn the most intense words than it was to learn the mildest words. For *S*s trained to say the most intense word, the ratio of learners to non-learners was six to one; for *S*s trained to say the mildest word, the ratio was two to one.

The 180 learners were assigned to groups as follows. For the *S*s trained to say the most intense word there were 40 *S*s in each of the three groups (Shout, Aloud, Whisper). For the *S*s trained to say the mildest word there were 20 *S*s in each of the three groups.

C. RESULTS

1. *Trials to Learn*

The various groups differed in the number of trials required to reach criterion in the learning series. These data are presented in Table 1. The most

TABLE 1
MEAN TRIALS TO LEARN

	Reinforced stimuli			
	Most intense		Mildest	
Groups	Men	Women	Men	Women
Whisper	67	39	59	58
Aloud	44	56	60	65
Shout	41	40	82	70

striking finding in this table is that *S*s learned to say the most intense words somewhat faster than the mildest words. There is also an interesting interaction between Groups and Reinforced Stimuli. The Shout group learned the most intense stimuli faster than the other two groups; but the Shout group learned the mildest stimuli slower than the other two groups. The results for the Whisper group are approximately opposite to those of the Shout group.

The significance of these trends was tested by an analysis of variance, which is presented in Table 2. (*Direction* refers to the intensity of the reinforced words, List 1 versus List 5.) The significant Direction F indicates that the most intense words were learned faster than the mildest words. The significant (.06 level) Groups × Direction F suggests that speed of learning in the Shout and Whisper groups does, as noted above, depend on whether most intense or least intense words are being learned. This finding holds more for men than for women; note in Table 1 that the women's Whisper group learned the most intense words faster than the other groups, whereas the men's Whisper group learned slower than the other two groups. This sex difference is reflected in the significant Direction × Sex interaction.

TABLE 2
ANALYSIS OF VARIANCE OF THE LEARNING DATA

Source of variation	*df*	MS	*F*	*P*
Groups	2	125		
Direction	1	12,591	11.2	.001
Sex	1	727		
Groups × Direction	2	3,283	2.9	.06
Groups × Sex	2	3,339	3.0	.05
Direction × Sex	1	387		
Groups × Direction × Sex	2	1,226		
Within	168	1,120		
Total	179			

2. *Stimulus Generalization*

On each trial of the generalization series *S* said one of four words. Since there were 80 trials, he made a total of 80 responses. His generalization gradient consists of the number of responses made to each of the four lists in the generalization series. The gradients for each group are presented in Figure 1. The gradients on the left side of the figure represent generalization from the most intense words (List 5) to less intense words; the gradients on the right side represent generalization from the least intense to more intense words.

Examination of Figure 1 reveals that *S*s did not simply transpose, i.e., respond only to the list adjacent to the reinforced list. Rather, they generalized. In the left hand gradients approximately half (40 out of 80) responses were to List 4, but the other half were to the remaining lists. In the right hand gradients more than half the responses were made to Lists 2, 3, and 4. For both directions of generalization there was a spread of response from the reinforced stimuli to stimuli distant on the intensity dimension. Thus this new paradigm was successful in producing stimulus generalization gradients.

Comparing the two sides of Figure 1, it is clear that the left-hand gradients are steeper than the right-hand gradients. Within each side there appear to be only very slight differences in the slopes of the gradients. Since the slopes are essentially linear, the measure used was the difference in response between extreme lists. This score was List 4 minus List 1 for the left side of Figure 1 and List 2 minus List 5 for the right side. An analysis of variance was performed on these slope scores. There was no significant difference between groups (Shout, Aloud, Whisper), between sexes, or in any of the interactions. The only significant variable was Direction: generalizing from intense to mild words yielded steeper slopes than generalizing from mild to intense words ($F = 4.66$, $p < .05$ for 1 and 168 *df*).

D. DISCUSSION

One purpose of this paper was to test the generality of previous findings. It was found that shouting facilitates the learning of intensely aggressive words and hinders the learning of mildly aggressive words; whispering had opposite effects. These results, which held for men more than for women, are consistent with the matching principle enunciated in previous studies: there is a tendency to match intensity of response to intensity of aggressive verbal stimuli.

The stimulus generalization data were not consistent with the matching

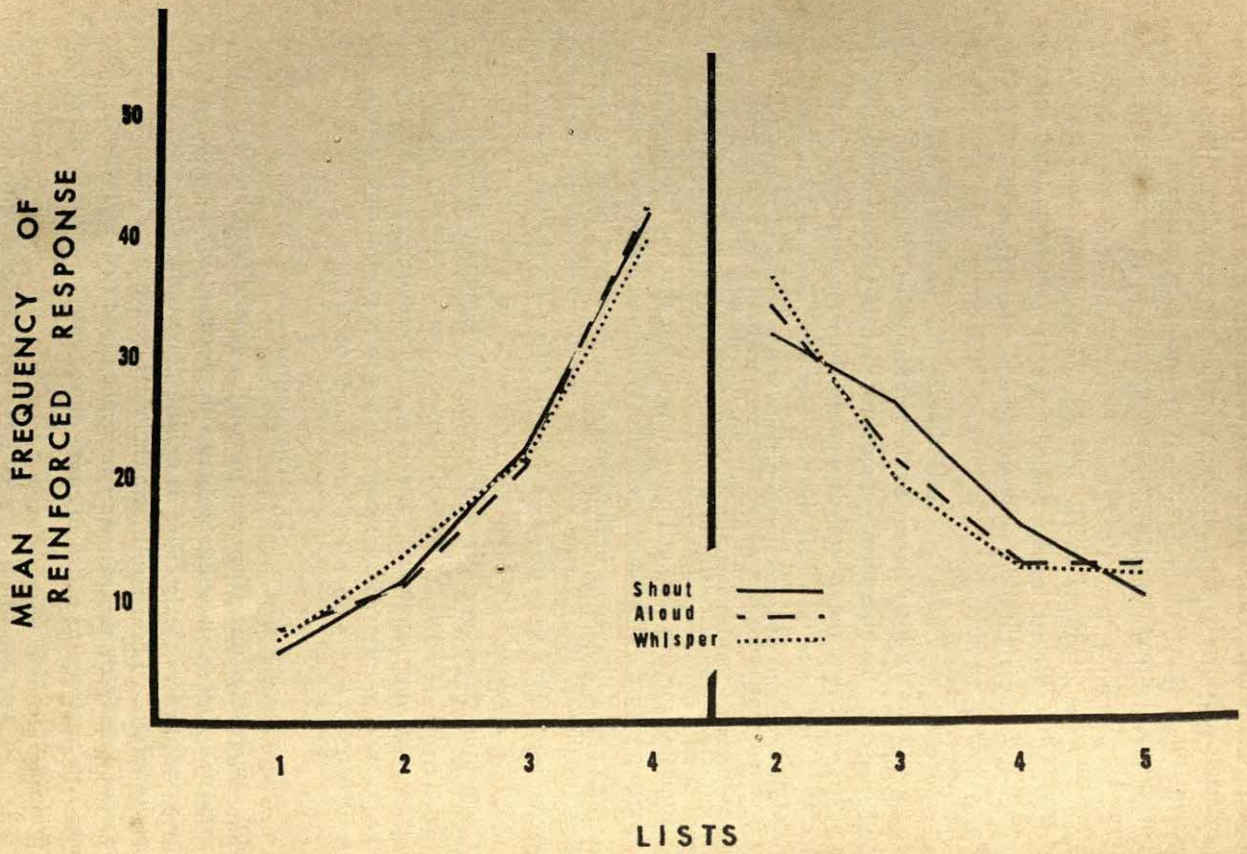


FIGURE 1
STIMULUS GENERALIZATION GRADIENTS

principle; no differences were found between the Whisper and Shout generalization gradients. This finding tends to limit the generality of the matching principle. Evidently, the matching principle applies only to the learning aspects of this new paradigm and not to stimulus generalization.

The effects of direction of generalization, found in previous studies, were confirmed in the present experiment. Generalizing from intense to mild aggressive stimuli yields a significantly steeper gradient than generalizing from mild to intense stimuli. The learning data are also consistent: faster learning on the intense end than on the mild end of the aggression dimension. Thus, it may be concluded that: (*a*) direction of generalization is a potent variable in this new paradigm; and (*b*) the matching principle receives equivocal support.

The second purpose of this study was to try a new method of studying stimulus generalization. *S* was presented with five words per trial, each of a different intensity of aggression. After being trained to say the extreme words (most intense or least intense), these reinforced words were omitted. It was found that there was a spread of response not only to adjacent words (close in intensity value) but also to distant words. The gradients resemble those derived from generally-accepted methods of studying stimulus generalization. Note that during learning *all* the words that subsequently appeared in the generalization series were negatively reinforced. In the absence of stimulus generalization there would have been an equal frequency of response to these words. The only plausible explanation of the obtained gradients is that there was a spread of response strength from the reinforced list of words to the others. The more similar the stimuli, the greater the frequency of response. As stimuli became progressively distant, response strength (frequency) diminished. The regularity of the relationship between distance (along the intensity dimension) and response strength suggests that this new paradigm may be fruitful in investigating stimulus generalization.

E. SUMMARY

A new stimulus generalization paradigm was tested. Five words of differing intensities of aggression were presented on each trial, and *S*s learned to say either the most or least intense word. Since different words were used on each trial, *S* had to respond to the intensity and not merely memorize words. Generalization was tested by omitting the reinforced class of words.

The following findings emerged: (*a*) there were regular gradients of stimulus generalization, suggesting the utility of the new paradigm; (*b*) the matching principle (tending to match response intensity to stimulus intensity) re-

ceived equivocal support; and (*c*) direction of generalization was, as in previous studies a potent variable; generalizing from intense to mild stimuli yielded steeper gradients than generalizing from mild to intense stimuli.

REFERENCES

1. BUSS, A. H. Stimulus generalization and aggressive stimuli. *J. Exper. Psychol.*, 1961, **61**, 469-473.
2. ———. The Psychology of Aggression. New York: Wiley, 1961, pp. 307.
3. GEER, J., & BUSS, A. H. Generalization of a nonverbal response to aggressive verbal stimuli. *J. Exper. Psychol.*, 1962, **63**, (in press).

Department of Psychology
University of Pittsburgh
Pittsburgh 13, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 417-424.

STIMULUS GENERALIZATION WITH AGGRESSIVE VERBAL STIMULI: A NEW PARADIGM*[1]

Department of Psychology, University of Pittsburgh

ARNOLD H. BUSS

A. INTRODUCTION

The purpose of this study was two-fold. First, it attempted to test the generality of previous findings concerning stimulus generalization with aggressive verbal stimuli. Second, it tested a new method for studying stimulus generalization.

A previous study (1) demonstrated that there is a tendency to match the intensity of a vocal response to the intensity of aggressive words. Direction of generalization was also found to be important. When generalizing from intense to mild aggressive stimuli, shouting the words steepens the generalization gradient and whispering flattens it. When generalizing from mild to intense aggressive stimuli, shouting flattens the gradient and whispering steepens it. These findings hold whether the response is vocal or motor (3). Will consistent results be found when the stimulus generalization paradigm is altered?

B. METHOD

1. *Stimuli*

The stimuli are words that have been scaled for intensity of aggression (2). They were divided into five lists of 20 each on the basis of intensity scale values. The range of scale values for the word groups was as follows:

1	1.8 - 3.5
2	4.0 - 4.5
3	5.0 - 5.5
4	5.7 - 6.7
5	7.0 - 8.8

Note that there is no overlap between lists. It was necessary to vary the range within lists in order to prevent overlap between lists. There are more words in the middle of the scale than at the ends, which means that the range of the middle groups must be smaller.

---

* Received in the Editorial Office on January 29, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported by Research Grant M-2861 from the National Institute of Mental Health, United States Public Health Service.

2. *Procedure*

The stimuli were presented five at a time in the learning series. The five words on any given trial consisted of one member from each of the five lists of words:

1	2	3	4	5
competitive	pest	anger	loathing	war

Since a learning series of 100 trials was necessary, words from each of the five lists were recombined, each word being used a total of five times. A given word never reappeared with its previous partners. Thus the next time *competitive* appeared, the other words on the trial were:

bruise browbeating cruel annihilation

The words were presented in random order on a given trial, and the sequence of presentation from trial to trial was randomized, with the limitation that no word be repeated until all the other members of its intensity group had been presented.

The words were presented by fives (one from each intensity list) on a memory drum. *S* was given the following instructions.

> I am going to show you a series of words. They will be presented five at a time, and each group of five words will appear for eight seconds. What you have to do is read *one* of the words aloud. Now you must make up your mind within eight seconds which word to read and then say it aloud. I will tell you whether you are right or wrong, and in that way you will get the hang of it.

Half the *S*s were reinforced for saying the most intense word; if they said one of the four milder alternatives, they were told *Wrong.* The other half were reinforced for saying the mildest word; if they said one of the four more intense words, they were told *Wrong.*

The criterion of learning was five consecutive correct responses. If *S* did not reach criterion by 100 trials, he was dropped from the experiment.

At the end of the learning series *S* was given the following instructions.

> Now we are going to switch over here (second memory drum), but I want you to keep going the way you have been. The only difference is that now there are four words in each group. I will no longer tell you whether you are right or wrong.

Then the generalization series commenced. This series differed from the learning series in two ways. First, there were 80 trials rather than 100. Second, there were only four words per trial. The words that had been reinforced in the learning series did not appear in the generalization series. *S*s who were reinforced for List 5 (most intense words) did not see these words in the generalization series. Similarly, *S*s who were reinforced for List 1 (mildest words) did not see these words in the generalization series.

In the absence of the words that he had been reinforced for saying, *S* was forced to choose one of the four remaining (previously unreinforced) words. He might simply transpose, saying only the next most intense word (List 4) or the next mildest word, depending upon whether List 1 or List 5 had been reinforced during learning. If this were to occur, there would be no generalization, merely transposition; the paradigm would be useless for investigating stimulus generalization. The success of the paradigm requires that there be a spread of response from the reinforced group of words to the other four groups.

3. *Design*

The first independent variable, direction of generalization, has already been mentioned. Half the *S*s generalized from intense to mild stimuli, and half generalized from mild to intense stimuli.

The second independent variable was intensity of the reinforced vocal response. One group of *S*s was instructed to shout the words and was given preliminary practice to insure shouting. One group was instructed to say the words aloud in a normal speaking voice. A third group was instructed to whisper the words. Thus the three groups were Shout, Aloud, and Whisper.

The last independent variable was sex of *S*. Half the *S*s were men, and half were women.

4. *Subjects*

*S*s were college students recruited from Introductory Psychology classes. A total of 235 *S*s took part in the experiment. Of these, 180 reached criterion, and 55 did not. There were no differences between the experimental groups or the sexes in the number of non-learners. However, direction of generalization was important, for it was easier to learn the most intense words than it was to learn the mildest words. For *S*s trained to say the most intense word, the ratio of learners to non-learners was six to one; for *S*s trained to say the mildest word, the ratio was two to one.

The 180 learners were assigned to groups as follows. For the *S*s trained to say the most intense word there were 40 *S*s in each of the three groups (Shout, Aloud, Whisper). For the *S*s trained to say the mildest word there were 20 *S*s in each of the three groups.

C. RESULTS

1. *Trials to Learn*

The various groups differed in the number of trials required to reach criterion in the learning series. These data are presented in Table 1. The most

TABLE 1
MEAN TRIALS TO LEARN

	Reinforced stimuli			
	Most intense		Mildest	
Groups	Men	Women	Men	Women
Whisper	67	39	59	58
Aloud	44	56	60	65
Shout	41	40	82	70

striking finding in this table is that *S*s learned to say the most intense words somewhat faster than the mildest words. There is also an interesting interaction between Groups and Reinforced Stimuli. The Shout group learned the most intense stimuli faster than the other two groups; but the Shout group learned the mildest stimuli slower than the other two groups. The results for the Whisper group are approximately opposite to those of the Shout group.

The significance of these trends was tested by an analysis of variance, which is presented in Table 2. (*Direction* refers to the intensity of the reinforced words, List 1 versus List 5.) The significant Direction F indicates that the most intense words were learned faster than the mildest words. The significant (.06 level) Groups × Direction F suggests that speed of learning in the Shout and Whisper groups does, as noted above, depend on whether most intense or least intense words are being learned. This finding holds more for men than for women; note in Table 1 that the women's Whisper group learned the most intense words faster than the other groups, whereas the men's Whisper group learned slower than the other two groups. This sex difference is reflected in the significant Direction × Sex interaction.

TABLE 2
ANALYSIS OF VARIANCE OF THE LEARNING DATA

Source of variation	*df*	MS	*F*	*P*
Groups	2	125		
Direction	1	12,591	11.2	.001
Sex	1	727		
Groups × Direction	2	3,283	2.9	.06
Groups × Sex	2	3,339	3.0	.05
Direction × Sex	1	387		
Groups × Direction × Sex	2	1,226		
Within	168	1,120		
Total	179			

2. *Stimulus Generalization*

On each trial of the generalization series *S* said one of four words. Since there were 80 trials, he made a total of 80 responses. His generalization gradient consists of the number of responses made to each of the four lists in the generalization series. The gradients for each group are presented in Figure 1. The gradients on the left side of the figure represent generalization from the most intense words (List 5) to less intense words; the gradients on the right side represent generalization from the least intense to more intense words.

Examination of Figure 1 reveals that *S*s did not simply transpose, i.e., respond only to the list adjacent to the reinforced list. Rather, they generalized. In the left hand gradients approximately half (40 out of 80) responses were to List 4, but the other half were to the remaining lists. In the right hand gradients more than half the responses were made to Lists 2, 3, and 4. For both directions of generalization there was a spread of response from the reinforced stimuli to stimuli distant on the intensity dimension. Thus this new paradigm was successful in producing stimulus generalization gradients.

Comparing the two sides of Figure 1, it is clear that the left-hand gradients are steeper than the right-hand gradients. Within each side there appear to be only very slight differences in the slopes of the gradients. Since the slopes are essentially linear, the measure used was the difference in response between extreme lists. This score was List 4 minus List 1 for the left side of Figure 1 and List 2 minus List 5 for the right side. An analysis of variance was performed on these slope scores. There was no significant difference between groups (Shout, Aloud, Whisper), between sexes, or in any of the interactions. The only significant variable was Direction: generalizing from intense to mild words yielded steeper slopes than generalizing from mild to intense words ($F = 4.66$, $p < .05$ for 1 and 168 *df*).

D. DISCUSSION

One purpose of this paper was to test the generality of previous findings. It was found that shouting facilitates the learning of intensely aggressive words and hinders the learning of mildly aggressive words; whispering had opposite effects. These results, which held for men more than for women, are consistent with the matching principle enunciated in previous studies: there is a tendency to match intensity of response to intensity of aggressive verbal stimuli.

The stimulus generalization data were not consistent with the matching

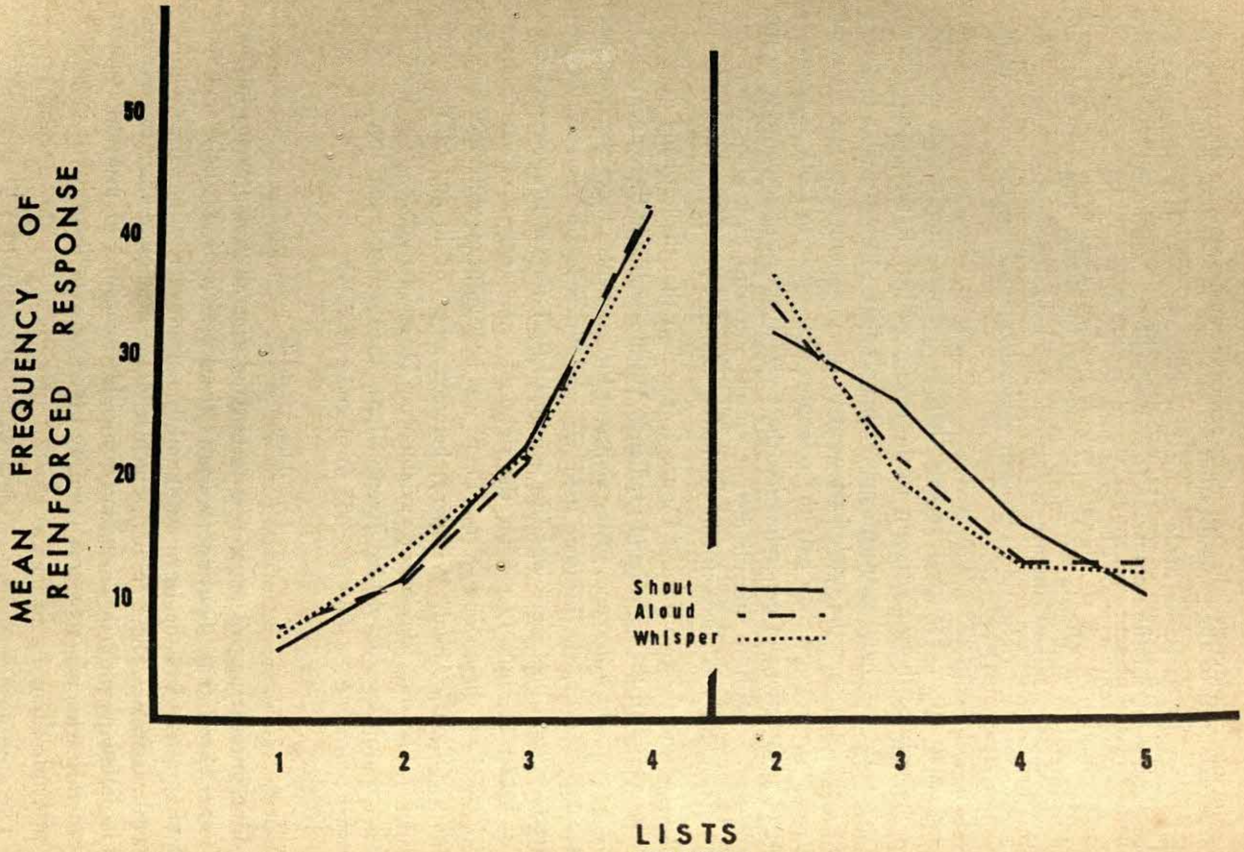


FIGURE 1
STIMULUS GENERALIZATION GRADIENTS

principle; no differences were found between the Whisper and Shout generalization gradients. This finding tends to limit the generality of the matching principle. Evidently, the matching principle applies only to the learning aspects of this new paradigm and not to stimulus generalization.

The effects of direction of generalization, found in previous studies, were confirmed in the present experiment. Generalizing from intense to mild aggressive stimuli yields a significantly steeper gradient than generalizing from mild to intense stimuli. The learning data are also consistent: faster learning on the intense end than on the mild end of the aggression dimension. Thus, it may be concluded that: (*a*) direction of generalization is a potent variable in this new paradigm; and (*b*) the matching principle receives equivocal support.

The second purpose of this study was to try a new method of studying stimulus generalization. *S* was presented with five words per trial, each of a different intensity of aggression. After being trained to say the extreme words (most intense or least intense), these reinforced words were omitted. It was found that there was a spread of response not only to adjacent words (close in intensity value) but also to distant words. The gradients resemble those derived from generally-accepted methods of studying stimulus generalization. Note that during learning *all* the words that subsequently appeared in the generalization series were negatively reinforced. In the absence of stimulus generalization there would have been an equal frequency of response to these words. The only plausible explanation of the obtained gradients is that there was a spread of response strength from the reinforced list of words to the others. The more similar the stimuli, the greater the frequency of response. As stimuli became progressively distant, response strength (frequency) diminished. The regularity of the relationship between distance (along the intensity dimension) and response strength suggests that this new paradigm may be fruitful in investigating stimulus generalization.

E. SUMMARY

A new stimulus generalization paradigm was tested. Five words of differing intensities of aggression were presented on each trial, and *S*s learned to say either the most or least intense word. Since different words were used on each trial, *S* had to respond to the intensity and not merely memorize words. Generalization was tested by omitting the reinforced class of words.

The following findings emerged: (*a*) there were regular gradients of stimulus generalization, suggesting the utility of the new paradigm; (*b*) the matching principle (tending to match response intensity to stimulus intensity) re-

ceived equivocal support; and (*c*) direction of generalization was, as in previous studies a potent variable; generalizing from intense to mild stimuli yielded steeper gradients than generalizing from mild to intense stimuli.

REFERENCES

1. BUSS, A. H. Stimulus generalization and aggressive stimuli. *J. Exper. Psychol.*, 1961, **61**, 469-473.
2. ————. The Psychology of Aggression. New York: Wiley, 1961, pp. 307.
3. GEER, J., & BUSS, A. H. Generalization of a nonverbal response to aggressive verbal stimuli. *J. Exper. Psychol.*, 1962, **63**, (in press).

Department of Psychology
University of Pittsburgh
Pittsburgh 13, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 425-440.

ANTICIPATORY BEHAVIOR: TEMPORAL MATCHING BY NORMAL AND PSYCHOTIC SUBJECTS*

University of Pittsburgh School of Medicine and Western Psychiatric Institute and Clinic

H. E. KING

A. INTRODUCTION

There are several sets of circumstances that render the human subject inefficient in the expected degree of control of his organized behavior or in the adequacy of his responsiveness to what is happening about him; uncommon stress, sleeplessness, drugged states, amentia and psychosis, to name but a few. In an effort to specify more precisely the nature of such conditions of maladaptation, both clinical (4, 10, 11, 27) and experimental writers (1, 19, 28, 29) have frequently mentioned defects in proper anticipation by their subjects or failures to show the expected qualities of adjustment in present behavior to what is yet-to-happen. Porteus (23, 25) and Robinson (26), in particular, have singled out errors in foresight, the ability to plan a course of action or in attitudes toward the future as being characteristic of the psychotic state itself, reactions which are said to become even more prominent following prefrontal lobotomy or other psychosurgical intervention (9, 12, 20, 22). The rather complex behavior observed and the methods employed in such investigations have usually stressed the total response of the individual to his situation and his organized ways of responding to the flux of potential stimuli about him. There is present, in the general statement, an implication for observation and test on a more limited and controlled scale, however, and we shall raise here one such question by a comparison of the responses of normal and schizophrenic persons to timed stimuli which are specially apt for the observation of anticipatory behavior. The term anticipation clearly implies action through time. It may be instructive to observe, therefore, the responses of a group of persons in so patent a state of maladaptiveness as schizophrenia to a situation in which timed behavior is brought into controlled and limited focus.

B. METHOD

As anticipatory behavior may be based on the extension in time of "systematizations" (6, 21) of either perceptual or action factors, it has

---

* Received in the Editorial Office on January 29, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

been thought best to reduce both elements to relatively simple terms in this experiment. The test procedure is based upon the ability of *S* to hear a succession of regularly recurring auditory stimuli, to fall in with the rhythmic "beat" by a finger-tapping accompaniment and to continue with the beat for a sufficiently extended period to permit a measure of the accuracy of his attempted synchronization. Several elementary rhythms, based on systematically differing interpulse intervals, have been explored in this way.

The selection of the intervals to be matched was made empirically, to provide a range from the "easy" to the "difficult" when rated subjectively by normal persons. The test stimuli, consisting of interpulse intervals of 1.0, 1.5, 2.0, 2.5 and 3.0 seconds, were selected to provide a series of five regular tempos—in steps of half a second—over a two second range.[1] A series of sharply defined, clearly audible clicks was presented as a train of regularly recurring sounds with which the subject was to try to synchronize his response. The clicks, which served to delimit the actual stimulus intervals, were produced electronically by means of a .001 second pulse fed to a loudspeaker placed in the midline, 18 inches in front of the subject and facing him. The loudness of the clicks, when measured at the distance of the average (subject) ear, was 54 decibels. Each click-producing electric pulse was also fed simultaneously to a marking pen that recorded a graphic trace of the stimulus sequence on a rapidly moving, continuous paper strip.[2]

S was instructed to listen to the audible "beat" for as long as he wished before attempting to tap along with it. Once he felt that "he had the feel of the tempo," he was to begin to tap the response key in such a way as to make his taps coincide as closely as possible with the succession of audible clicks. A demonstration run, with a stimulus interval of 0.5 seconds, was performed by the experimenter. *S* was then invited to attempt to synchronize his taps with a stimulus sequence in a like manner. The point of the task was grasped easily by all *S*s and little need arose to extend or to repeat these simple verbal instructions. Each *S* was asked to continue to respond, once he had begun, until told to stop—despite any seeming errors in coincidence, for these were to be expected and are made, to some extent, by everyone. Once a stimulus sequence was begun, and *S* had launched his attempt at synchronous response, thirteen responses were recorded. The performance on the first three trials was universally discarded, as a familiarization period for all *S*s, and the remaining ten trials were then scored as the measure of matched-interval performance for that stimulus sequence. Following a rest interval, and reassuring comment offered by the experimenter, the procedure was repeated for the next stimulus interval to be matched. Matching was always begun with the 1.0 second stimulus interval ("easy") and then lengthened by 0.5 second steps in successive stimulus sequences to the ("difficult") 3.0 second stimulus interval.

---

[1] There is evidence, from the work of other investigators, that intervals of more than one second involve a "memory factor" (3) and that such intervals are clearly separable from the signals which delimit them, forming stimulations . . . not belonging to the same "psychological present" . . . but a distance between a past and present event (8, p. 117).

[2] The use of notched paper travel and monitored power supply to the kymograph motors, in conjunction with the accuracy made possible by electronic production of the pulses delimiting stimulus interval length, maintained the error in the recording system at less than 1 per cent when calibrated against external standards.

Each response of *S* was written, by a signal marking pen, adjacent to and on the same kymographic strip that was recording the stimulus sequence. This system permits the observation and measurement of the degree of simultaneity between each audible click and the response made by *S* and the direction of response errors in time, if any. The actual matching required is, of course, to the interval between audible clicks, the clicks themselves serving only to delimit the interval to be matched. Since it is *S*'s task to anticipate just when the next click in a sequence will arrive (in time) and to organize his response to coincide with it, perfect response would be exactly coincident and of zero displacement in time. All responses by *S* that came before the succeeding click in a stimulus sequence, and which therefore precede it, will be called (—) in this paper, while those which follow it will carry no coded prefix. The parentheses are used as a reminder that we are coding early response as minus time for convenience and that that the number score is not negative in the mathematical sense.

S was seated, alone, in a sound-deadened chamber throughout, making his responses by means of a telegraph key placed in the midline before him on a table 25″ in height. The response key was mounted on rubber cushioning; the displacement for full closure being 1 mm. at a tension of 200 gms. The experimenter and all recording apparatus were placed in an adjoining room, in voice contact with the subject by intercom. The experimenter entered the test chamber during the instruction period, of course, and again during each rest period for comment and conversation with *S*. Before each sitting, and again at its end, a series of simultaneous electrical calibrating pulses was routinely fed to the marking pens to maintain maximal accuracy in monitoring the alignment and operation of the test equipment itself.

C. SUBJECTS

The accuracy with which a series of selected temporal intervals could be matched was observed for two groups of subjects:

1. *Normal Group*

A sample of 30 non-patients, 16 male and 14 female, was observed. The age range extended from 21 to 63 years, with an average age of 26.1 years. These *S*s were members of the local housekeeping, business, maintenance and professional staff who were taking part in a broad series of psychological studies being made at the same time (17). All were self-supporting members of the working community, non-volunteer and without history of any psychiatric care.

2. *Schizophrenic Group*

A sample of 30 chronic schizophrenic patients was also observed in matched-interval performance. Fifteen of the patients were male and 15 female. The age range was from 27 to 58 years, with an average age of 47.6 years. All of these patients had been committed to institutional care and bore psychiatric diagnoses of long standing.[3] Their chronicity is attested by an average duration of illness, since first committed hospitalization, of 16.6 years; no patient was included in the group with less than three years of illness since his first committed hospitalization. None of the patients was receiving tranquilizing medication at the time of testing, nor had any received such medication for a minimal period of two months before taking part in this experiment. These patients, also, were participating in a broader series of psychological studies being made at the same time (17), as had the normal group. The difficulty met in completing all of the tests in this program has, in all likelihood, exercised a selective effect upon the makeup of our patient group by eliminating the more disturbed patients from inclusion. This probable influence is borne out by the evidence of ancillary behavior ratings (13), psychomotor evaluations (13, 31) and Porteus Qualitative Scores (24) that are available for each individual in the sample. When judged by these criteria, the group selected for matched-interval study falls near the middle of the range of disordered behavior usually found among chronic schizophrenic populations (13, 16, 24).

D. RESULTS

It must be made clear at the outset that it is the *S*'s estimate of the interval

---

[3] The hospital diagnoses of the patient group:

	n
Schizophrenic Reaction, Chronic Undifferentiated Type	8
Schizophrenic Reaction, Paranoid Type	12
Schizophrenic Reaction, Catatonic Type	7
Schizophrenic Reaction, Simple Type	2
Schizophrenic Reaction, Hebephrenic Type	1
	30

between clicks that is being observed and that this is the actual stimulus under study. The clicks themselves serve only to delimit the intervals and as reference points from which to measure the accuracy of interval estimation. If an interval is judged to be matched by *S* (subjective time) before it is actually over (chronometric time), his response will come before the next click in a sequence. An error of this kind is scored as (—), and is measured in milliseconds of deviation from coincidence.

1. *Normal Group*

The performance of the normal group, in its attempt to match the intervals set by the experimenter, is summarized by the mean error scores given in Table 1. These are averaged scores for all trials of all *S*s for a particular interval length; a measure of constant error for the total group. It is notable that all of the scores are minus and that the mean response of normal *S*s, therefore, consistently came before the succeeding click in any sequence, no matter what the length of the stimulus interval under test. The same data have been represented in Figure 1, where the duration of each stimulus interval has been drawn on a scale proportional to its length and the degree of error in matching performance is plotted on the same scale. These data, on the whole, indicate a degree of success at approximating the various interval lengths by the normal group (over-all error *circa* 16 per cent), marked by a consistent tendency for mean response to precede the end of the stimulus interval. They further indicate a relatively more accurate matching, both in terms of the constant error score and its standard deviation, for the 1.5 and 2.0 second intervals and show a closer approximation of response to actual stimulus duration for intervals toward the middle of the test series than was shown at either extreme; 1.0 or 3.0 seconds.

TABLE 1
MEAN ERROR SCORES AND STANDARD DEVIATIONS FOR MATCHED-INTERVAL PERFORMANCE BY THE NORMAL AND SCHIZOPHRENIC GROUPS (IN SECONDS)

Stimulus intervals (in seconds)	Normal group (n = 30)		Schizophrenic group (n = 30)		
	M	σ	*M*	σ	*t*
1.0	(—).308	.43	(—).117	.49	1.58
1.5	(—).140	.39	.365	.77	3.29**
2.0	(—).145	.40	.682	1.25	3.40**
2.5	(—).359	.79	.284	1.45	2.10*
3.0	(—).558	.96	.246	2.20	1.80

* Significant at .05 level
** Significant at .01 level

As it may also be useful to evaluate performance of this type in terms of the absolute error, disregarding sign and focusing rather on the amount of deviation apart from its direction, the responses of the normal group have been scored in this manner as well. The absolute error scores for the normal group are given in Table 2 and are presented graphically in Figure 2. An inspection of these data reveals a regular increment as we pass from the shorter to the longer stimulus intervals, with regularly increasing standard deviations. A line of best fit, drawn among the points representing averaged absolute error scores plotted in Figure 2, describes this regularity as being nearly linear in form.

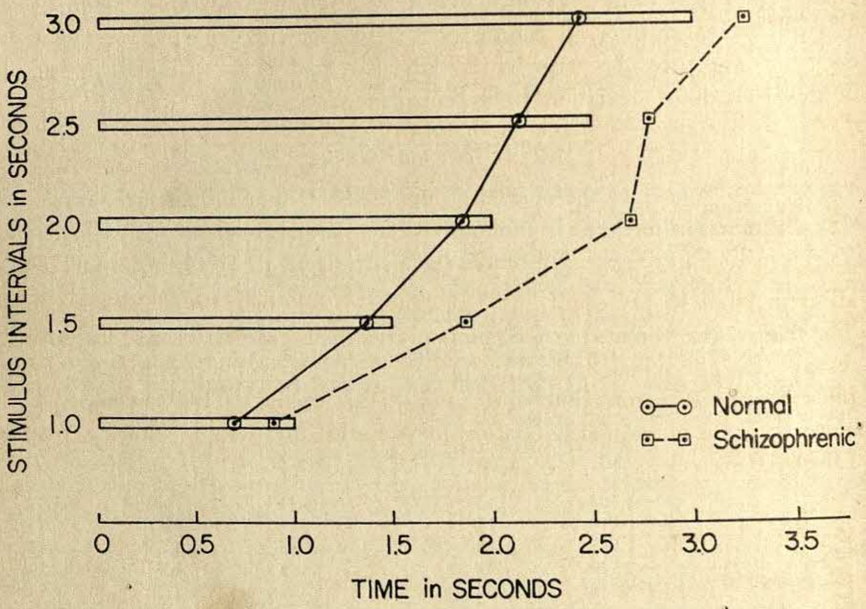


FIGURE 1
INTERVAL-MATCHING PERFORMANCE BY THE NORMAL AND SCHIZOPHRENIC GROUPS; MEAN ERROR SCORES

2. *Schizophrenic Group*

The mean error scores achieved by the group of schizophrenic *S*s, the constant error terms, are also summarized in Table 1. It is notable that only one value is minus (—), for the 1.0 second interval, and that the mean response of schizophrenic *S*s for all other intervals observed falls after the succeeding click in a stimulus sequence. A graphic display of these averaged responses is shown in Figure 1. The schizophrenic group was also able to

TABLE 2
Absolute Error Scores and Standard Deviations for Matched-Interval Performance by the Normal and Schizophrenic Groups (in Seconds)

Stimulus intervals (in seconds)	Normal group (n = 30)		Schizophrenic group (n = 30)		t
	M	σ	M	σ	
1.0	.570	.24	.809	.47	2.44*
1.5	.717	.25	1.160	.47	4.45**
2.0	1.077	.33	1.747	.57	5.47**
2.5	1.232	.36	2.240	1.03	8.00**
3.0	1.474	.51	2.575	1.30	4.21**

* Significant at .05 level.
** Significant at .01 level.

achieve a rough approximation of the various actual interval lengths (overall error *circa* 18 per cent), but tends most of its errors on the positive side. The dashed line connecting the points plotted in Figure 1, for ease in reading, reveals that the schizophrenic group was relatively less accurate in matching the 1.5 and 2.0 second intervals, toward the middle of the test series, and more accurate in its approximation of stimulus durations at either extreme; 1.0 or 3.0 seconds.

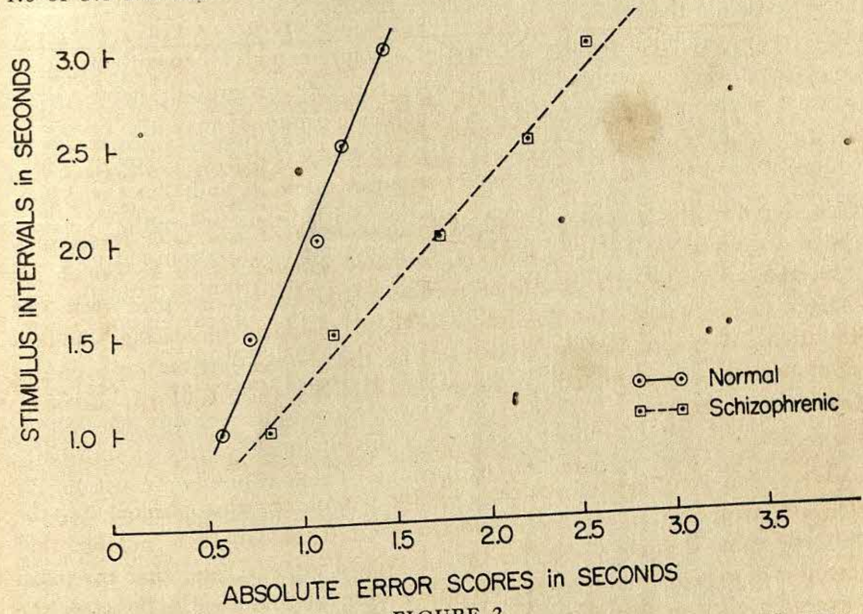


FIGURE 2
Interval-Matching Performance by the Normal and Schizophrenic Groups; Absolute Error Scores

The absolute error scores of the schizophrenic group, shown in Table 2 and Figure 2, follow a regular increase as we pass from the shorter to the longer stimulus intervals, together with corresponding regular increases in group standard deviations. A line of best fit, drawn in Figure 2 among the points representing averaged absolute error values for the group, describes this regularity as being nearly linear in form.

3. *Intergroup Comparison and Subanalyses*

It is apparent from an inspection of the curves presented in Figures 1 and 2, and by the *t* values determined between the group means given in Tables 1 and 2, that the normal and schizophrenic groups differ markedly in the form of their response to the task of matching intervals. The response of the normal group consistently precedes the succeeding click in a sequence, while the mean response of the patient group follows in performance on four of the five stimulus sequences. Three of these five values attain the level of statistical significance and the remaining two, while not achieving this level, are consistent with the trend established. Neither group shows a tendency toward increased constant error with progressive lengthening of the stimulus interval; the normal group mean errors being greater at the extremes of the series while the errors of the schizophrenic *S*s are greatest in mid-series. Both groups exhibit a regular increase of absolute error values with progressively lengthened stimulus interval, with the rate of growth being more apparent in the performance of the schizophrenic group. It may also be noted that the standard deviation values for the constant error scores show a slight tending toward increase for the longer stimulus intervals and that the schizophrenic group regularly increases variation with interval length. Neither group exhibits a strong trend toward related increase of variation and interval length for its absolute error scores, although a consistent hint is present.

a. Inter- and Intra-Interval Correlations. A series of correlations, prepared for each group separately, on the relation of matched interval performance between intervals of differing length revealed only low order correspondence, (*circa* .36), even between adjacent steps of 0.5 seconds separation. As a low task-reliability might account for this finding, a series of split-half correlations was also prepared for each group. These values are consistently high, (*circa* .87), indicating a substantial reliability in performance within a given test interval.

b. Age and Sex. Sample correlations and subgroup intercomparisons were performed to examine for a tendency to sex- or age-related response, as these descriptive factors were similar but not identical for the two *S* groups. No systematic or significant relationship appeared.

c. Time of Response. Although it is difficult to be certain about what would constitute a relevant measure of such a general factor, correlations were computed between performance on interval-matching and lift reaction time measures, both simple prepared lift and serial lift time to auditory stimuli, which were available for the same *S* groups (17). No systematic or significant relationship to either appeared.[4]

E. DISCUSSION

The experimental findings reflect a distinctly different pattern of matched-interval responding by the normal and schizophrenic *S*s in their attempt to "project" short stimulus intervals and to synchronize their behavior with the simplified temporal series presented them. The tendency of the normal group to respond early, so to speak, and for the schizophrenic *S*s to respond late seems unmistakable, as does the trend for schizophrenic *S*s to show, regularly, increased variability in performance and a greater magnitude of absolute error on all samples of interval-matching observed. An evaluation of the possible meaning of the differences observed between groups must at first carry us somewhat afield, into comment upon the possible relation of interval-matching to the processes of sensory psychophysics, time judgment and "voluntary conditioning," however, before we may reasonably expect to relate the data of this experiment to the more general questions of anticipatory behavior that have been raised.

Not much is known about the form that the responses of normal *S*s might be expected to take on tasks of this kind. Upon first consideration the task might seem to resemble, and therefore be related to, the process of time estimation, as it has been classically judged and measured (5). Although the details of such a possible relationship remain to be explored, it appears, upon closer inspection, that studies made of the judgment of (short) intervals of time are more momentary and center more exclusively upon sensory factors, employ a quite different subject-set and frequently involve verbal estimates of clock-time in a way quite different from the sensori-motor, repetitive set of interval-matching which makes no demand upon individual verbalized concepts of actual or clock-time (32). Of the several standard methods for experimentally evaluating time judgment (5), only the Method of Reproduction, in which *S* tries to duplicate in action an interval set by an action of the examiner, bears much apparent relation to the effort to synchronize a motor response with an anticipated sensory event. Bartlett and Bartlett (3), in reporting the results of their recent experiments which are similar to those described here but are addressed primarily to very short intervals ($<$ 1.0 seconds), have, rather, likened the experimental set for response synchrony to one of repeated reaction time with the uncertainty element reduced to a minimum. Testing for synchronized finger response to auditory stimuli in a time range of 0.125 to 0.50 seconds, these investigators found the errors of their (normal) *S*s to be predominantly those of striking the response key slightly early, behavior which is in close accord with that shown to characterize interval-matching by our own normal group. Although the literature of classical time judgment is in a state of some confusion, owing to the use of a variety of methods, interval lengths and points of reference, there is general agreement that very short intervals tend to be consistently overestimated (8, 30, 32). For this reason, as well as for the differences in task that have been described, the data generated by a set for interval-matching, or response synchrony, would seem to reflect a different order of behavior, to be considered in its own right rather than regarded as an aspect of classical time judgment. There can be little doubt that both concern the use made by the individual of the perception of time passage in the more general sense, but they deviate enough in the method of appraisal and the relation of time appreciation to other facets of behavior to produce a differing order of response.

---

[4] The fact that individual responses were regularly made either before or after the succeeding click indicates that these responses are not merely reaction times to the clicks themselves, as does the fact that the size of the error differs on different trials and for different intervals. Deliberate "follow the click" instructions to another group of subjects produced performance which yields a different order of error scores, similar error scores for all test intervals and performance which showed substantial and positive correlation with the ancillary measures of simple and serial lift reaction time (17).

The consistent underestimation of intervals reported here, and by Bartlett and Bartlett (3), may be considered to be reminiscent of the negative time-error of psychophysics, as the context requires a sequential procedure with a stimulus being followed by a judgment of the stimulus. An attempted match to a prior sensation often has been likened to a match of the present instant with a fading trace; diminishing either toward zero or toward the mean of all sensation (33). Such a match may, thus, fall short in relative magnitude when it is compared with the physical dimensions of the original stimulus. The absolute error term would also be expected to increase, generally, with the magnitude of the stimulus in such situations. Although the diminished matches given by our normal *S*s, and their linear increase in absolute error with interval length, do fit these features of the negative time-error description, the available data do not permit further examination of the possibility at the moment. An analogy might also be raised to the formation of "voluntary conditioned responses," and what has been observed in normal interval-matching might be regarded as a response moving forward in time as an interval length became sensed or conditioned. On descriptive grounds this seems somewhat unlikely, for the instructions do not prescribe voluntary response to the succeeding click in a sequence, which might then typically advance in time as it was learned, but prescribe, instead, a direct response to the interval itself. It should be noted, however, that the direction and relative ease and immediacy with which this phenomenon is produced does conform with what has been described as typical acquisition for this type of response among normal subjects (18).

For whatever reasons the form of normal response in interval-matching tends consistently toward early response, it has done so clearly for the present group of *S*s and for those reported in generally comparable experimental contexts (3, 7). This is, in itself, an interesting finding deserving further investigation, and it is under probing study at the present time. Obviously, more extended observations will need to be made, and more analytic questions be experimentally phrased, to resolve the basis for the phenomenon. The contrast of the moment, however, lies more directly between the performance of the sample normal and schizophrenic groups and their differing modes of response to an identical task. In making this comparison the performance of the normal group will be taken as the standard, as a typical control reference group, rather than assuming a zero value based upon physical time, against which to measure relative performance by the schizophrenic *S*s.

When matched-interval performance is related directly to a clock-time standard, neither group achieves true coincidence and each appears to err about equally (15-20 per cent), but to fall on opposite sides of the coincident (zero) point. Taking the responses of the normal *S*s as a control or standard group against which to compare patient achievement, however, indicates a clear and consistent retardation in performance by the schizophrenic *S*s. The slowness of the patient group in synchronizing response with the rhythmic temporal stimuli has been shown to be not the simple consequent of a shift in set to respond to the succeeding click in a sequence, which might be thought to have occurred as an "easy way out" of compliance with the more difficult instructed set to respond to interval length. Prior study of the establishing of voluntary conditioned responses among such populations would appear to rule out this mechanism also, as a possible basis upon which to account for the present differences in group performance, since identical processes and latencies have been found to characterize the acquisition of VCR's both for normal *S*s and psychiatric patients (18).

The systematic differences observed between the matched-interval responses of normal and schizophrenic *S*s may rest in something not more specifically delimitable, at the moment, than an inability to adapt in present behavior to that which is yet to happen, or, again, may have at its root some better defined characteristic which is generally retarded in the behavior of the schizophrenic patient. One possible explanation is offered by the far-reaching psychomotor inadequacy and lowered response strength typical of patients with severe disorders in patterns of organized behavior, whether these clinically detectable disruptions of behavior are of physiological, psychiatric or neurological origin (13, 15, 16). Assuming the presence and involvement of this often reported slowness of psychomotor responsivity among our patient sample, one might expect to find a lengthened latency and an increased variation of response made to a "concept of an interval," even though perceptual systematization of the stimulus was formed and retained just as it is by the normal individual. It is far from agreed that this kind of recurring stimulus and response can be so readily fractionated, however, the continuing adjustment and modifying flow of *S* and *R* being what it is for the usual acquisition of sensorimotor skills (2). It might also be expected, were this the situation, that a positive correlation would express itself between the scores on interval-matching performance and the ancillary measures of simple and serial lift reaction times available for the identical S group, which was not observed. Taking the broader view of what is meant by the term adaptive behavior, it might also be reasoned, since the instructed goal is one of synchrony in behavior, that such a hypothetically increased response latency would be taken into account in the adjustment of the total organism to the task, just as a compensation may be made for defects in vision or in the muscular control of an action. Despite these seeming contradictions, the general retardation of psychomotor responsiveness in schizophrenia appears to offer the most satisfactory temporary hypothesis as to the mechanism concerned in producing the consistently positive constant errors in the interval-matching performance by the patients relative to normal response. It is altogether possible, in view of what is known about the way in which this characteristic manifests itself (13, 14, 16, 17, 31), that interval-matching behavior may sample a relatively independent aspect of the general state and might, therefore, express only nonsignificant correlations with other psychomotor measures, such as the lift reaction time scores obtained in this experiment (13, pp. 13-15, 58, 78, 101).

One of the basic similarities in performance by the two groups, rather than a difference between them, should also be commented upon briefly for what it may imply about the psychological basis of interval-matching. The finding, in each group, of high intra-interval correlations, together with low correlations between scores for matching responses on intervals of different lengths, suggests the presence of a form of psychometric function in the performance of this task. It would appear that a direct stimulus extension upon a single physical continuum (clock-time) may tap several different levels of ability, even over a rather narrow range in time (1-4 seconds). This finding may relate to the comment made by Clausen, in discussing data derived by the Method of Reproduction, the only classical Time Judgment procedure to bear a descriptive similarity to interval-matching: "The method of Reproduction produces judgments closer to the stimulus interval than [the other methods] but shows more instability in the intercorrelations between the subtests . . . this indicates that Reproduction measures a function of less uniform character" (5, p. 759). The intervals in question were 5, 10 and 15 seconds in length and produced data of a quite different order than did the methods of Verbal and Operative interval estimation. "Reproduction yields results that are not interchangeable with results from the other methods [of time judgment]" (5, p. 760). No immediate explanation is offered or is required, at the moment, for the relatively slower performance given by the patients on the 2.0 second interval. This stimulus length may differ psychometrically from the others in the series or may possibly be more subject to influence by "waves of attention" as these have been discussed by Woodworth (33). Until the finding is replicated it may be taken to be a chance occurrence, which, while statistically significant and fitting the general trend of the data from other interval lengths, may nonetheless be characteristic only of the present sampling in its magnitude.

Whatever may be its cause, a different and apparently defective degree of adequacy in temporal projection seems to characterize the performance of patients in the schizophrenic group, leading, in this experimental context, to faulty anticipatory behavior in matching short intervals presented in regular stimulus patterns. Neither age nor sex differences in the composition of the sample groups observed account for the characteristically positive constant error of schizophrenic performance. This contrast in performance may rest upon a retardation in the general responsiveness of schizophrenic *S*s or, again, be a more basic indication of a derangement of the over-all adjustment and anticipatory function by patients in the sample group. It would be quite unsafe to generalize from the results of normal-patient contrast in performance in so restricted a situation to the broader fabric of complex and organized human behavior, especially as anticipation probably normally proceeds in a continuing flow of interplay between receptors and effectors, rather than in the more stilted framework of this experiment where a given response can in no way affect the succeeding stimulus (2, 6). It may be useful to note, however, that these observations, made in a more limited context but drawn from a broad statement about the quality of human behavior in disrupted states (1, 4, 10, 11, 19, 27, 28, 29), has revealed clear and systematic differences in normal and schizophrenic responsive actions. That anticipatory behavior, as it may be observed and measured in terms of temporal matching, is not equivalent for the normal and psychotic *S*.

F. SUMMARY

1. A statement is often made that defective states of the human organism such as conditions of stress, sleeplessness, amentia, psychosis or intoxication, are characterized by a faulty anticipation of the future or by errors in relating present behavior to what has yet-to-happen. Clinical and experimental writers have both commented upon defects in foresight or in attitudes toward the future in chronic schizophrenia; defects said to be exacerbated by psychosurgical intervention.

2. This experiment has attempted to test, by means of observations made in a more limited context, the ability of both normal and schizophrenic *S*s to project a regular temporal stimulus series and to synchronize behavior to make a coincident response to these stimuli. Regular stimulus intervals of 1.0, 1.5, 2.0, 2.5 and 3.0 seconds have been presented and the accuracy of interval-matching response has been measured for sample groups of normal and schizophrenic *S*s.

3. Two clear and consistent trends emerged from an analysis of the data: (*a*) normal *S*s tended to respond early, before the stimulus was at an end (negative constant error); and (*b*) schizophrenic *S*s tended to respond late, after the stimulus interval was at an end (positive constant error).

The differences in mean performance between the groups were statistically significant for most test intervals, with conforming trends where group performance did not differ significantly. The responses made by each sample have also been analyzed in terms of absolute error, group variation and by correlations made between and within intervals, as well as with ancillary data available on the sample characteristics of age, sex and lift reaction time.

4. (*a*) The consistently early responding by the normal *S*s has been discussed in terms of its possible relation to the negative time-error, and the findings of classical time judgment and voluntary conditioning. (*b*) the sys-

tematic differences in the performance of normal and schizophrenic *S*s have been discussed and a possible mechanism sought in terms of a shift in set, the establishment of voluntary conditioned responses and the characteristic psychomotor retardation of patients with a severe behavior disorder. The latter appears to offer the most adequate hypothesis of the moment.

5. Whatever the basis for the phenomena observed, different and seemingly defective performance was obtained when observations were made, drawn from a broad statement about human behavior in defective states, in the more limited context of sensory anticipation and temporal matching.

REFERENCES

1. BARACH, A., & KAGAN, L. Disorders of mental functioning produced by varying the oxygen tension of the atmosphere. *Psychosom. Med.,* 1940, **2**, 53-67.
2. BARTLETT, F. C. Anticipation in human performance, pp. 1-17 in Essays in Psychology dedicated to David Katz. Uppsala: Almqvist and Wiksells, 1951.
3. BARTLETT, N., & BARTLETT, C. Synchronization of a motor response with an anticipatory sensory event. *Psychol. Rev.,* 1959, **66**, 203-218.
4. BROWER, J., & BROWER, D. The relation between temporal judgment and social competence in the feebleminded. *Amer. J. Ment. Def.,* 1947, **51**, 619-623.
5. CLAUSEN, J. An evaluation of experimental methods of time judgment. *J. Exp. Psychol.,* 1950, **40**, 756-761.
6. DODGE, R. Anticipatory reaction. *Science,* 1933, **78**, 197-203.
7. DUNLAP, K. Reactions to rhythmic stimuli, with attempts to synchronize. *Psychol. Rev.,* 1910, **17**, 399-416.
8. FRAISSE, P. Psychologie du Temps. Paris: Presses Universitaires, 1957.
9. FREEMAN, W., & WATTS, J. Psychosurgery. Springfield, Ill.: Thomas, 1942.
10. HOCH, P. Pharmacologically induced psychoses. Ch. 84 *In* Arieti, S. (*Ed.*), American Handbook of Psychiatry, II, N.Y.: Basic Books, 1959.
11. ISRAELI, N. The psychopathology of time. *Psychol. Rev.,* 1932, **39**, 486-491.
12. JONES, R. Personality changes in psychotics following prefrontal lobotomy. *J. Abn. & Soc. Psychol.,* 1949, **44**, 315-328.
13. KING, H. E. Psychomotor Aspects of Mental Disease. Cambridge: Harvard Univ. Press, 1954.
14. ———. Psychomotor tests. Ch. 13 *In* Zubin, J. *et al. Experimental & Abnormal Psychology.* N.Y.: Columbia Univ. Bkstr., 1957.
15. ———. Non-specific indicators of CNS damage. *Fed. Proc.,* 1961, **20**, 329.
16. ———. Some explorations in psychomotility. In *Psychiatric Research Reports,* #14. Washington: Amer. Psychiatr. Ass'n, 1961.
17. ———. Sensory and motor performance as a function of behavior disorder (in preparation).
18. KING, H. E., & LANDIS, C. A comparison of eyelid responses conditioned with reflex and voluntary reinforcement in normal individuals and in psychiatric patients. *J. Exp. Psychol.,* 1943, **33**, 210-220.
19. LAGARZA, C., & WORCHEL, P. Time and space orientation in schizophrenics. *J. Abn. & Soc. Psychol.,* 1956, **52**, 191-194.
20. LEBEAU, J. Psycho-chirurgie et Fonctions Mentales. Paris: Masson, 1954.
21. LINDNER, R. An experimental study of anticipation. *Amer J. Psychol.,* 1938, **51**, 253-266.

22. PETRIE, A. Personality and the Frontal Lobes. London: Kegan Paul, 1952.
23. PORTEUS, S. The Porteus Maze Test and Intelligence. Palo Alto: Pacific, 1950.
24. ———. The Maze Test: Recent Advances. Palo Alto: Pacific, 1955.
25. PORTEUS, S., & PETERS, H. Maze test validation and psychosurgery. *Genet. Psychol. Monogr.*, 1947, **36**, 3-86.
26. ROBINSON, M. Psychosurgery and the Self. N.Y.: Grune, 1954.
27. SCHILDER, P. Psychopathology of time. *J. Nerv. & Ment. Dis.*, 1936, **83**, 530-540.
28. Tufts College Institute of Applied Experimental Psychology. Handbook of Human Engineering Data, 2nd Ed., Special Devices Center, Office of Naval Research, 1951, Part VII.
29. WALLACE, M. Future time perspective in schizophrenia. *J. Abn. & Soc. Psychol.*, 1956, **52**, 240-245.
30. WALLACE, M., & RABIN, A. Temporal experience. *Psychol. Bull.*, 1960, **57**, 213-236.
31. WEAVER, L. Psychomotor performance of clinically differentiated chronic schizophrenics. *Perc. & Motor Skills*, 1961, **12**, 27-33.
32. WOODROW, H. Time perception. Ch. 32 *In* S. Stevens (*Ed.*), *Handbook of Experimental Psychology.* N.Y.: Wiley, 1951.
33. WOODWORTH, R. Experimental Psychology. N.Y.: Holt, 1938.

University of Pittsburgh
School of Medicine
Western Psychiatric Institute & Clinic
3811 O'Hara Street
Pittsburgh 13, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 441-452.

ABSTRACT CONDITIONING*

Department of Psychology, University of Kentucky

JERRY E. WATERS AND FRANK KODMAN, JR.

A. INTRODUCTION

A new method of conditioning, called abstract conditioning, was introduced by Welch (12) and Corn-Becker, Welch and Fisichelli (3). Following Pavlovian (11) and Hullian conditioning (7), Welch originally hypothesized the process of abstract conditioning as one means of explaining the development or enhancement of suggestibility during an induction of hypnosis. Corn-Becker, Welch and Fisichelli describe abstract conditioning as a type of generalized conditioning resulting from "conditioning factors of past associations" (3, p. 221). The nature of abstract conditioning may be shown by an example from the Corn-Becker, Welch and Fisichelli experiment. Subjects were told that they were to be used in an experiment on palmar skin resistance, and GSR electrodes were attached to their hands. Stimulus words were projected one at a time on a screen. For example, the word *green* was shown, and immediately thereafter a green light was turned on. The word *music* was projected, and music was introduced.. Seven words were presented in this manner, each paired with an event or object which it symbolized. On the eighth, twelfth, or sixteenth trial the word *electric shock* appeared. This stimulus was not reinforced. In a group of naive subjects, 73.0 per cent gave galvanic skin responses to this stimulus which were either greater or of longer duration than the galvanic skin response given to any of the other stimuli. In a control group in which no verbal stimuli were reinforced, the corresponding percentage of 13.0 was reported. One might argue that the results were due to the subject's expectation of a shock to his hand, but there was no correlation between the subjects' statements of their expectation or nonexpectation and the size and duration of the galvanic response.

Corn-Becker, Welch and Fisichelli called attention to the similarities between the abstract conditioning process, Pavlovian conditioning and the induction techniques of the hypnotist. They maintain that in reinforcing each word with its physical actuality, the action of the hypnotist and the experi-

---

* Received in the Editorial Office on January 29, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

menter in the conditioning situation are duplicated. According to Corn-Becker, Welch and Fisichelli, "The hypnotist, in telling the subject that his eyes are tired (conditioned stimulus equals 'your eyes are tired'), produces a situation in which the subject's eyes will become tired (unconditioned stimulus equals staring at a bright object resulting in a conditioned response of tired eyes to the words), and the conditioning experimenter sounded a bell (conditioned stimulus), followed by food (unconditioned stimulus) and the conditioned response of salivation to bell" (3, p. 213). This view which emphasizes the influence of spoken language on bodily control appears to be in essential agreement with F. A. Allport, who stated, "Perfectly normal individuals also show at times an immediate and deliberated response to commands. These effects are based on deeply fixed habits of association between word sounds and the bodily movements they signify" (1, p. 243).

There has been some evidence given on the relationship between simple conditioning and suggestibility as manifested in the hypnotic situation. Das (4) reports a correlation of .51 between "depth of hypnosis" and conditioning of the eye-blink. The work of Hudgins (6), Cason (2) and Menzies (8) also offer support for a conditioning process as the underlying mechanism of suggestion and hypnosis.

Pattie (10) in a critique of abstract conditioning pointed out that the question of the existence of this form of conditioning should be further investigated and also its relationship to hypnotizability, since "Neither Welch nor anyone else has tested the hypnotic susceptibility of any of the subjects used in the conditioning experiment" (10, p. 12). Following Pattie's former suggestion, the primary objective of this study is to provide additional support for the validation of abstract conditioning. Also, in an effort to extend these findings as well as those of Corn-Becker, Welch and Fisichelli, subjects are classified as either suggestible or nonsuggestible.

A convenient test of suggestibility is the Release Test. Eysenck and Furneau (5) found a correlation of .64 between responses to this test and hypnotizability. Susceptibility to hypnosis was determined by the responses to 21 suggestions, which were rated on three-point scales ranging from no observable response to a complete response.

B. METHOD

1. *Suggestibility Test*

A total of 118 freshmen college women were administered the Release Test. The Release Test apparatus consisted of a pliable rubber ball filled with water and a glass cylinder which was connected to the ball with plastic tubing. A scale with millimeter units showed the changes in hand pressure as *S* compressed and released the rubber ball and thus varied the water level in the glass cylinder.

Each *S* began the test with the rubber ball compressed to a pressure level arbitrarily referred to as "zero pressure level." *S* was instructed to "maintain constant pressure on the ball" while listening to the recorded voice of the experimenter. Three experimental test trials, each of 105 seconds duration, were given with 60-second rest periods between trials. During the experimental trials *S* listened to the following tape-recorded suggestions.

> Now I want you to concentrate on the rubber ball you are holding in your hand. You have been doing your best to keep constant pressure on the ball. However, your hand is not as strong as it was at first. You feel your grip relaxing. You can feel the arm and hand muscles becoming more and more relaxed—more and more relaxed.
>
> Your hand continues to relax and as the muscles relax, your hand becomes weaker and weaker and weaker. This weakness is causing your fingers to relax, and you are losing your grip on the rubber ball.
>
> Now your fingers are opening wider and wider and wider. You can no longer keep a constant pressure on the ball and must relax your grip. You must relax your grip. You can feel the rubber ball expanding and expanding and expanding as you gradually loosen your grip. Your hand is very, very tired now, and you must release the ball. As you relax your hand you continue to release the ball, releasing more and more and more.
>
> The ball is expanding bigger and bigger and bigger as your hand is relaxing, relaxing, relaxing. Now your muscles are so very tired that you can no longer hold the ball but must completely release it. You must completely release it.

Initially, as a control trial, *S* attempted to maintain constant hand pressure for 105 seconds without suggestions of release. The four test periods (the initial observation period plus the three experimental test periods) were evaluated to determine each *S*'s group classification. In order to be classified as suggestible, *S* had to demonstrate a release score at the completion of the second experimental test period which exceeded her release score in the observation period by a minimum of 20 mm. Also, the 20 mm. increase in "release" had to be maintained in the third test period. The criteria adopted for a classification of nonsuggestibility were that: (*a*) no release score for any test period would exceed 20 mm.; and (*b*) no variability would exceed 10 mm. between the release scores for the four test periods. The measurements were made in units of 2 mm.

2. *Abstract Conditioning*

In the second session of this investigation, 20 *S*s who had been classified as suggestible and 20 *S*s who had been classified as nonsuggestible were randomly divided into two suggestible subgroups and two nonsuggestible subgroups. Following a similar procedure as that used in the Corn-Becker, Welch and Fisichelli experiment, each *S* was exposed to a conditioning situation involving auditory and visual stimulus words paired with specific events related to the meanings of the particular stimulus words. For example, the word *red* (conditioned stimulus) was presented and followed in close proximity by a red light (unconditioned stimulus). After nine stimulus words—*bell, green, music, flicker, red, blue, static, yellow, buzzer*—and their respective referents constituting reinforcement had been presented, a critical stimulus word *electric shock* was presented without reinforcement. The galvanic skin response (GSR) was used as the measure of conditioning.

The apparatus used in the conditioning procedure consisted of the electronic circuits and accessory devices for presenting the stimuli and the instrument, which was constructed by Nichols and Daroge (9) for measuring the GSR. In brief, this instrument is designed so that changes in *S*'s resistance produce changes in the potential across the *S* and can be read as percentage of change in skin resistance.

The nine reinforcing stimuli were presented automatically by means of two electronic interval timers and a relay system which were operated from a room adjoining the testing room in order to limit extraneous auditory stimuli. Six-volt AC bulbs provided the appropriate visual stimuli for the subjects while a six-volt doorbell and doorbuzzer provided the "bell" and "buzzer" auditory stimuli. Recorded harmonica music was presented from a tape recorder. Static was produced by an ordinary radio tuned to a frequency which always produced static noise. Stimulus words were presented by means of an Argus slide projector, model P.A. 200. Each word appeared on a projection screen for a total of 14 seconds as in the Corn-Becker, Welch and Fisichelli study.

Under reinforcement conditions the appropriate visual or auditory stimulus occurred during the last four seconds of a 14-second interval—the four seconds of reinforcement overlapping with the four seconds of the word presentation. Intervals between words varied in multiples of five seconds and ranged from 20 to 60 seconds according to a predetermined random order. Each *S* served as her own control, receiving stimulus words under reinforcement and nonreinforcement (control conditions) in which words were presented without an accompanying stimulus. Order effects were counterbalanced in the design by presenting reinforced stimulus words initially to one-half of the *S*s while the remainder received nonreinforced stimulus words first.

Therefore, the stimuli were presented so that differences in groups (suggestible versus nonsuggestible), orders (reinforcement first versus reinforcement second), words (mean of the preceding nine stimulus words versus the word *electric shock*) and reinforcement (reinforcement versus nonreinforcement) could be tested by an analysis of variance design. This design permits partition of the experiment into four conditions as follows: (*a*) Reinforcement first: reinforcement stimulus words presentd before non-reinforced stimulus words. (*b*) Reinforcement second: reinforced stimulus words presented after nonreinforced stimulus words. (*c*) Nonreinforcement first (control): nonreinforced stimulus words presented before reinforced stimulus words. (*d*) Nonreinforcement second: nonreinforced stimulus words presented after reinforced stimulus words.

C. RESULTS

Eight of 20 suggestible *S*s completed a final observation period following the third experimental release-test period. The final observation period began 60 seconds following the third experimental test period and was conducted in the same manner as the initial observation period. If release scores were the result of a fatigue effect, the scores in the final observation period should have equalled or exceeded release scores in the third experimental test period. However, release scores for the eight *S*s completing the final observation periods were of lesser magnitude than release scores for the same *S*s obtained in the third experimental test periods. With a t value of 3.28, release score differences between the final observation period and the third experimental test period were significant beyond the .01 level. Furthermore, a t value 1.64 ($p = .20$) for the eight *S*s indicated that release scores were not significantly different in the initial and final observation periods; therefore, a fatigue effect was ruled out.

1. *GSR Measurements*

In the conditioning situation, a percentage ohmic change was obtained for each *S* for each stimulus word. However, in order to meet some basic assumptions involved in producing valid analyses of the results, a transformation of the data into log conductance change units was required. In accordance with the suggestions of Nichols and Daroge, all GSR scores referred to in this investigation are the logarithms to the base ten of the conductance change in micromhos plus one micromho × one hundred.

2. *Analysis of Variance*

Mean log conductance change (hereafter called mean LCC) measures for group, order, reinforcement and word factors appear in Table 1. Mixed analysis of variance design of the GSR data as shown in Table 2, second and third rows, indicated no significant differences between suggestible and nonsuggestible groups or the orders of presentations of the stimulus words.

TABLE 1
MEAN LOG CONDUCTANCE CHANGE FOR GROUP, ORDER, REINFORCEMENT AND WORD FACTORS

Factors	Subfactors	Mean
Groups	Suggestible	89.70
	Nonsuggestible	91.61
Order	Order 1	93.04
	Order 2	88.28
Reinforcement	Reinforcement	110.71
	Nonreinforcement	70.60
Words	Mean of nine stimulus words	65.25
	Electric shock	116.05

TABLE 2
MIXED ANALYSIS OF VARIANCE DESIGN OF GSR DATA

Scores	*df*	Sum of squares	Mean squares	*F*	*P*
Between *S*s	39				
Groups (R)	1	146	146	—	
Order (A)	1	907	907	—	
AR	1	522	522	—	
Error b	36	217,672	6,046.44		
Within *S*s	120				
Reinforcement (B)	1	64,360	64,360	73.62	.001
Words (C)	1	103,276	103,276	118.14	.001
BC	1	19,648	19,648	22.48	.001
BR	1	34	34	—	
CR	1	48	48	—	
BCR	1	945	945	1.08	>.20
AB	1	8,688	8,688	9.94	.005
AC	1	1,306	1,306	1.49	>.20
ABC	1	149	149	—	
ABR	1	158	158	—	
ACR	1	4	4	—	
ABCR	1	290	290	—	
Pooled *S*s	108	94,415	874.21		

As indicated in Table 1, the mean LCC of 110.71 for reinforcement exceeded the mean LCC of 70.60 for nonreinforcement and the mean LCC of 116.05 for the word *electric shock* was larger than the mean LCC of 65.25 for the mean of the nine stimulus words. These differences were significant at the .001 level as shown by the analysis of variance in Table 2. Furthermore, the words × reinforcement interaction was statistically significant at the .001 level. The means plotted in Figure 1 indicate that significantly greater GSR differences occurred between the mean of the nine stimulus words and the word *electric shock* under experimental conditions of reinforcement than under conditions of nonreinforcement. This finding plus the significant differences in the reinforcement and word factors support the major hypothesis of this experiment concerning evidence for the existence of the phenomenon of abstract conditioning.

The GSR data plotted in Figure 2 represents the order × reinforcement interaction found to be significant at the .005 level. These data show that the mean differences between the two reinforcement and the two nonreinforcement conditions contributed to the interaction. However, the latter observation which suggests that prior reinforcement affected the following nonreinforced word presentations is of most concern here. A *t* value of 3.17, significant beyond the .01 level between the stimulus words' mean and the word *electric shock* in the nonreinforcement second condition indicated that abstract conditioning did not extinguish immediately but extended into the nonreinforced presentations. This latter measurement of conditioning was made nine minutes after the original abstract conditioning.

3. *Further Statistical Analyses*

A significant GSR difference between the mean of the nine stimulus words and the word *electric shock* could have occurred independently of the condition of reinforcement. To rule out this possibility, the data of the nonreinforcement first condition (control) were analyzed by a *t* test. Under the latter conditions, a *t* value of 1.99 resulted from a comparison of the stimulus words' mean and the final stimulus word *electric shock* and indicated that the difference failed to reach the .05 level of significance. Thus, as predicted, this finding indicated nonsignificant differences between the mean of the nine stimulus words and the word *electric shock* under control conditions. However, a *t* test computed independently for suggestible *S*s tested in the nonreinforcement first condition showed a *t* value of 2.37 significant at the .05 level and indicated that a significantly greater GSR response occurred for the word *electric shock* than for the stimulus words in

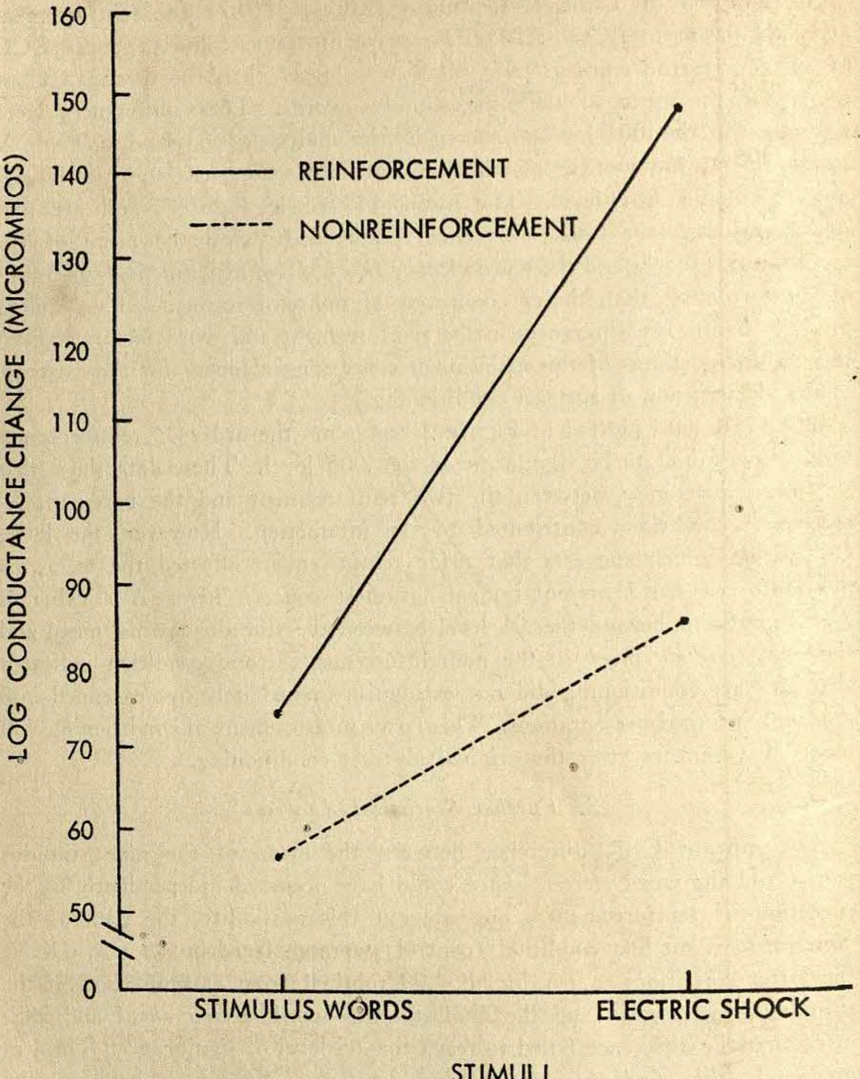


FIGURE 1
Log Conductance Change for the Mean of the Nine Stimulus Words and the Word Electric Shock Under Conditions of Reinforcement and Nonreinforcement

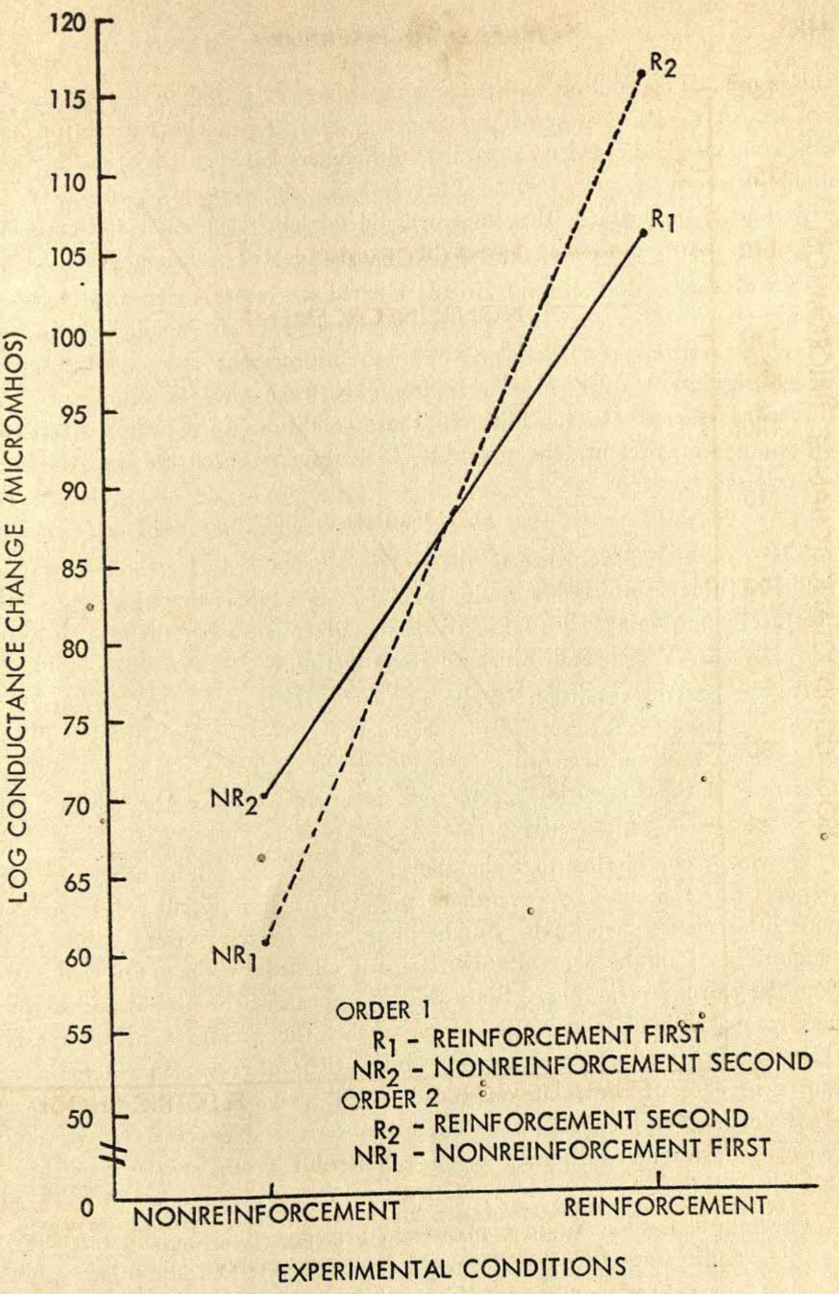


FIGURE 2
Log Conductance Change for Orders of Reinforcement First and Reinforcement Second Under Conditions of Reinforcement and Nonreinforcement

this group. This finding will be explained more fully in the discussion of the results. A *t* value computed for nonsuggestible *S*s in the nonreinforcement first condition indicated no significant differences between the stimulus words and the word *electric shock.* Therefore, in the control condition of the experiment, the critical stimulus word did not elicit a greater GSR response than the mean of the other stimulus words for the nonsuggestible *S*s.

None of the nine stimulus words elicited a greater mean GSR response than the word *electric shock* in the reinforcement first condition, the reinforcement second condition or the nonreinforcement second condition in which significant differences were found between the stimulus words and the word *electric shock.* Thus, in those conditions in which abstract conditioning was present, the greatest GSR response occurred for the word *electric shock.*

D. Discussion

The conditioned response to the unreinforced word *electric shock,* which followed nine reinforced stimulus words adds support for the existence of abstract conditioning. This finding is in agreement with the results of the Corn-Becker, Welch and Fisichelli investigation.

In the control condition of the present investigation, the suggestible *S*s gave a greater GSR response to the word *electric shock* than to the mean of the other stimulus words. Therefore, the occurrence of abstract conditioning in the suggestible group was inconclusive since the word *electric shock* elicited a significantly greater GSR response.

However, considering that the above *S*s were suggestible, a greater GSR response for the word *electric shock* might have been predicted even under nonreinforcement conditions. Furthermore, the largest mean GSR for the suggestible *S*s in the control condition was elicited by the word *bell* rather than the word *electric shock.* Since the GSR mean of the nine stimulus words was compared with the word *electric shock,* the greater GSR reaction to the word *bell* was obscured. These observations offer some explanation for the occurrence of the relatively greater GSR response to the word *electric shock* for suggestible *Ss* in the control condition. Nevertheless, the occurrence of abstract conditioning in the suggestible group was not verified in this investigation.

The nonsuggestible *S*s, in contrast to the suggestible *S*s, did not show a significant difference between the word *electric shock* and the other stimulus words under control conditions. Therefore, we would argue that differences between the critical word and the other stimulus words were due to the reinforcement factor. This finding supported the occurrence of abstract conditioning for the nonsuggestible *S*s.

The measurement of conditioning in the nonreinforcement second condition, which followed the experimental abstract conditioning session, indicated that conditioning did not extinguish immediately. A conditioned response to the word *electric shock* was demonstrated at least nine minutes after conditioning had been introduced. Further investigation will be necessary to determine the duration of abstract conditioning.

Although Welch in formulating the theory of abstract conditioning was mainly concerned with hypnosis, the title of his paper, *A Behavioristic Explanation of the Mechanism of Suggestion and Hypnosis,* implies that he was also concerned with the suggestion variable. In any event, as reported earlier, Eysenck and Furneau found a correlation of .64 between responses on the Release Test and hypnotizability. Also, physiological factors which both Welch and Corn-Becker, Welch and Fisichelli associated with the abstract conditioning process are present in the Release Test situation. For example, suggestions of "relaxation," "weakness," etc., are reinforced with true physiological fatigue resulting from the concentration and muscle tension required for maintaining constant pressure on the rubber ball. It is interesting to note that one suggestible *S* actually went into a hypnotic trance while listening to the recorded suggestions. It seems that suggestible *S*s would be more likely to exhibit abstract conditioning than nonsuggestible *S*s. However, the results indicated nonsignificant differences between suggestible and nonsuggestible groups in this respect.

In conclusion, it should be pointed out that neither Welch nor Corn-Becker, Welch and Fisichelli attempted to relate suggestibility as measured by the Release Test to abstract conditioning. Since the Release Test was the only measure used in diffierentiating suggestible and nonsuggestible subjects, the negative results of this study concerning a relationship between abstract conditioning and suggestibility do not support a rejection of Welch's contentions. Nevertheless, an explanation of the mechanisms underlying suggestion and hypnosis based on the occurrence of abstract conditioning is open to question. Certainly, further evidence is needed to substantiate the relationship between abstract conditioning and suggestibility.

E. SUMMARY

In order to investigate the phenomena of abstract conditioning, 20 suggestible and 20 nonsuggestible female *S*s were exposed to a conditioning situation involving auditory and visual stimulus words and reinforcing events

related to the meanings of the specific words. Although the results were inconclusive for suggestible *S*s, abstract conditioning was demonstrated for nonsuggestible *S*s by the occurrence of a greater GSR response to a critical stimulus word *electric shock.* However, the results failed to show a relationship between abstract conditioning and suggestibility as measured by the Release Test.

REFERENCES

1. ALLPORT, F. A. Social Psychology. Boston: Houghton-Mifflin, 1924.
2. CASON, H. The conditioned pupillary reaction. *J. Exp. Psychol.,* 1922, **5**, 108-146.
3. CORN-BECKER, F., WELCH, L., & FISICHELLI, V. Conditioning factors underlying hypnosis. *J. Abnorm. & Soc. Psychol.,* 1949, **44**, 212-222.
4. DAS, J. P. Conditioning and hypnosis. *J. Exp. Psychol.,* 1958, **56**, 110-113.
5. EYSENCK, H. J., & FURNEAU, W. D. Primary and secondary suggestibility: an experimental and statistical study. *J. Exp. Psychol.,* 1945, **39**, 485-503.
6. HUDGINS, C. V. Conditioning and the voluntary control of the pupillary light reflex. *J. Gen. Psychol.,* 1933, **8**, 3-51.
7. HULL, C. L. Hypnosis and Suggestibility. New York: Appleton-Century, 1933.
8. MENZIES, R. Further studies in conditioned vasomotor responses in human subjects. *J. Exp. Psychol.,* 1941, **28**, 457-482.
9. NICHOLS, R. C., & DAROGE, T. An electronic circuit for the measurement of the galvanic skin response. *Amer. J. Psychol.,* 1955, **68**, 455-461.
10. PATTIE, F. A. Theories of hypnosis. *In* R. M. Dorcus (*Ed.*), *Hypnosis and its Therapeutic Application.* New York: McGraw-Hill, 1956, 1-30.
11. PAVLOV, I. P. Conditioned Reflexes. London: Oxford Univ. Press, 1927.
12. WELCH, L. A behavioristic explanation of the mechanism of suggestion and hypnosis. *J. Abnorm. & Soc. Psychol.,* 1947, **42**, 359-364.

Department of Psychology
University of Kentucky
Lexington, Kentucky

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 453-455.

STEREOTYPES OF THE SCIENTIST AS SEEN WITH THE GORDON PERSONAL PROFILE AND GORDON PERSONAL INVENTORY*

Department of Psychology, University of Bridgeport

JOHN R. BRAUN

A. INTRODUCTION

Bendig and Hountras (1) investigated the stereotypes of the personality characteristics of research scientists held by University of Pittsburgh students. *S*s had to indicate whether typical members of various paired occupations (including research scientists, lawyers, businessmen, and engineers) would be equal or unequal on a number of characteristics. They found that the research scientist, as compared with the other three occupations was regarded as ". . . more intellectual, logical, orderly, persistent, studious, thorough, and also as being less charming, friendly, humorous, poised, and self-confident."

The present investigation dealt with the stereotypes of the personality characteristics of scientists held by Texas Christian University students in 1961. Instead of *S*s making direct comparisons using trait labels, they responded to personality inventories in terms of their conceptions of how a typical scientist or typical business executive would answer.

B. METHOD

*S*s were 42 male and female undergraduates in a child psychology class. Inventories used were the Gordon Personal Profile (GPP) and Gordon Personal Inventory (GPI).

Twenty-eight *S*s took the GPP (2) under two sets. Under one set they answered as they thought a scientist would answer, and under the other set as would a business executive.

Fourteen *S*s took the GPI (3) under these two sets. For both inventories the order of administration of the sets was counterbalanced. Significance of difference between the scores resulting from the two sets was evaluated by Wilcoxon paired-replicates tests.

---

* Received in the Editorial Office on January 31, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

C. Results and Discussion

Tables 1 and 2 present the means and standard deviations of inventory scores produced by the sets. With the GPP, the business executive set resulted in significantly higher scores on Ascendancy and Sociability, while the scientist set resulted in significantly higher scores on Responsibility and Emotional Stability. The business executive set also resulted in significantly higher Total scores, which would be the extent to which complimentary rather than derogatory alternatives were chosen.

TABLE 1
Gordon Personal Profile Means and Standard Deviations for Scientist and Business Executive Sets
(N = 28)

Variable	Scientist Mean	Scientist SD	Business Executive Mean	Business Executive SD
Ascendancy	—0.82	4.95	11.96**	3.24
Responsibility	13.43	3.99	6.32**	5.11
Emotional stability	6.29	5.41	0.57**	4.55
Sociability	—3.25	3.65	10.43**	3.65
Total	14.86	15.89	30.00**	8.72

** = Significantly different from Scientist Mean at .01 level

TABLE 2
Gordon Personal Inventory Means and Standard Deviations for Scientist and Business Executives Sets
(N = 14)

Variable	Scientist Mean	Scientist SD	Business executive Mean	Business executive SD
Cautiousness	9.93	4.22	6.71	4.53
Original thinking	13.00	2.85	1.57**	6.70
Personal relations	9.57	3.31	—0.21*	4.11
Vigor	8.36	4.27	8.21	6.13
Total	30.86	7.77	16.29**	17.09

* = Significantly different from Scientist Mean at .05 level.
** = Significantly different from Scientist Mean at .01 level.

With the GPI, the scientist set resulted in significantly higher scores on Original Thinking and Personal Relations. There were no significant differences between the sets on Cautiousness and Vigor. This time Total score was significantly higher for the scientist set.

Based on these findings we may say that business executives were seen as quite ascendant and sociable, rather vigorous and energetic though cautious, and moderately responsible. They were seen as having much worry and tension, not particularly concerned with ideas, and not having a great deal of faith or trust in people.

Scientists were seen as quite responsible, cautious, and concerned with ideas, rather vigorous and energetic, and rather high on degree of faith and trust in people. They were seen as having only an average amount of worry and tension and as being not at all ascendant and sociable.

In spite of differences in region, time, and technique of investigation, the results of the present study are consistent with those of Bendig and Hountras. The scientist is seen as high on intellectuality and low on sociability. Since adolescents place great stress on popularity it is easy to see why many may be reluctant to pursue a career in science. They see a scientist as one who lacks interest and skill in dealing with other people. Research into the most effective ways of modifying this stereotype is sorely needed.

D. Summary

In 1961, 42 Texas Christian University students took either the Gordon Personal Profile or Gordon Personal Inventory under sets to answer as would a typical scientist or a typical business executive. The resulting stereotype of the scientist was found to be quite similar to that found by Bendig and Hountras in 1958 with University of Pittsburgh students.

References

1. Bendig, A. W., & Hountras, P. T. College student stereotypes of the personality traits of research scientists. *J. Educ. Psychol.,* 1958, **49**, 309-314.
2. Gordon, L. V. Manual, Gordon Personal Profile. Yonkers, New York: World Book, 1953.
3. ———. Manual, Gordon Personal Inventory. Yonkers, New York: World Book, 1956.

Department of Psychology
University of Bridgeport
Bridgeport, Connecticut

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 457-563.

THE INFLUENCE OF "FAKE" PERSONALITY EVALUATIONS ON SELF DESCRIPTION*[1]

Department of Psychology and Education, Swarthmore College

GLORIA CAREY EVANS

A. INTRODUCTION

Personality tests in one form or another are currently being taken by a fairly wide range of individuals. The tests may be administered in psychological clinics, in industrial situations, in schools, and may be found in nonstandardized form in newspapers and magazines. It is not clear, however, what effect the results of these tests have on the individual and his subsequent behavior. Of course the test results and an interpretation of them are not always available to the individual, but the findings are frequently made known either in a direct or an indirect fashion to the testee. For example, a therapist may respond to his patient on the basis of a personality diagnosis or discuss the findings and their implications with him, the magazine reader is given a classificatory scheme for interpreting his score, and the employee who is not hired or promoted may infer that his test responses reflected on him adversely.

When test data are made available or the results are at least inferred, the individual may react in a number of ways. He may deny that the evaluation is valid, he may repress the findings, or he may attempt to modify himself and his behavior in the light of the evaluation. Such modifications might take the form of enhancing those characteristics which he deems desirable and lessening those which he feels are socially or personally undesirable.

One way of determining whether or not an individual attempts to modify his behavior is to note his performance on post-evaluation personality inventories. Obviously, an evaluation might influence behavior in ways other than verbal ones, but if verbal responses are modified it could be assumed that the evaluation had had some effect on the individual, and the direction of this effect could be observed.

In the present study, a personality inventory was administered and then

---

* Received in the Editorial Office on January 31, 1962 and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to express appreciation to Paul Meisel and the members of her experimental psychology class for their assistance in the execution of this study.

a supposedly genuine, but actually fake scoring of the inventory was returned before a second inventory was administered. Fake scorings were used in order that evaluation would be objectively the same for each subject (*S*) and also because they provided an opportunity to verify the findings concerning the fallacy of personal validation. Forer (1) and Sundberg (3) reported that college students tended to endorse fake personality evaluations as being descriptive of themselves, particularly if the evaluations contained vague, favorable statements. The present study employed extremely negative as well as positive personality ratings and attempted to make them specific by using graphic rating scales with the ends of the scales clearly labelled.

Thus, the aims of this study were: (*a*) to determine the effect of personality evaluations on future self descriptions; (*b*) to re-check the acceptability of fake personality evaluations; and (*c*) to determine whether negative as well as positive evaluations may be accepted and acted upon.

B. METHOD

1. *Design*

*S*s were asked to complete a personality inventory consisting of 225 items. Each inventory was numbered and *S*s were instructed to remember this number because personal names would be removed from the blanks before scoring. These inventories were taken away, completed, and returned individually by the *S*s. About a week later all of the *S*s reported to the same room to complete a second inventory, but it was explained that because of space limitations those *S*s with code numbers less than 50 were arbitrarily asked to report to another room. The *S*s remaining in the room (experimental group) were given the fake results of the inventory and then asked to complete a second inventory of 225 items. Those *S*s in the other room (control group) were given the second inventory without an intervening evaluation.

2. *Subjects*

*S*s were members of an introductory psychology course and participated in the study in order to receive part of the experimental credit required of all students.[2] They were aware that the course instructor was not involved in the study and that their performance would in no way affect their course standing. They were told that the three inventory administrators, who were students in an advanced course in psychology, were attempting to determine group norms for the college. As soon as the second inventory was completed, all *S*s were informed that the evaluations were bogus and a report on the study was later distributed to them.

---

[2] Unfortunately, anonymity was so carefully preserved that it was impossible to classify *S*s according to sex. The numbers of males and females, however, were judged to be approximately equal.

3. *Materials*

The first part of each of the two personality inventories consisted of 75 statements chosen at random from a variety of standard inventories and were to be answered "yes," "no," and "?". The second part of each inventory consisted of 150 adjectives which were to be answered on a four-point scale according to the degree to which each was characteristic of *S*. Although the first 75 statements were different for the two inventories, the same 150 adjectives were repeated on both forms.

The evaluations for the experimental *S*s were given out according to code numbers as a sort of after-thought or "in case you are interested" manner by the administrators. Since the *S*s were spread throughout the room it was not possible for them to see that each was receiving the same evaluation. On the evaluation sheet were ten graphic rating scales, on each of which a check mark was placed to indicate how *S* compared with other students from the college. The ten scales were labelled *anxiety, sociability, self-regard, organization, stability, inhibition, aggression, consistency, prejudice,* and *integrity*. In addition to marking off percentile points on each scale, the ends of the scales were labelled with descriptive terms taken from the list of adjectives; e.g., the sociability scale was labelled *withdrawn, cold* at one end and *outgoing, warm* at the other. Two of the scales (*organization* and *consistency*) were regarded as "fillers" and check marks close to the 50th percentile were placed on them. On the remaining eight scales the check marks were placed so that half of them were between the 75th and 100th percentiles and half were between the 0 and 25th percentiles.

C. RESULTS

The responses to the first 75 items on each inventory were not evaluated. These items were used in order to detract attention from the self-descriptive adjectives—the responses to which constituted the critical part of the study. It was felt that if the two inventories had been completely identical, *S*s might have been overly set to respond primarily on the basis of their recall of original responses.

1. *Responses to Adjectives*

Nine students from an advanced psychology course served as judges in order to select all the adjectives from the inventory which they felt were related to the eight critical scales included in the evaluations given to the experimental *S*s. Sixty-five adjectives were thus selected from the original list of 150 and the direction of the anticipated change was predicted. A comparison was then made between the two responses given by each *S*s to each of the 65 adjectives, with the direction of the change on the four-point scale being scored algebraically. Thus, if *S* rated the word *timid* as four, or extremely characteristic of himself the first time, and as three, or occasionally characteristic of himself the second time, the change was scored as +1 since the fake evaluation had indicated that *S* was highly aggressive and the prediction was that he would subsequently rate himself as less timid.

The experimental *S*s showed a significantly greater change in the predicted direction than did the control *S*s ($t = 4.34$, $df = 33$, $p < .001$). In fact, the control *S*s showed a small mean change in the opposite direction. Twenty-four of the 34 experimental *S*s changed in the predicted direction, seven showed no change, and three changed in the opposite direction. By comparison, 15 of the 34 control *S*s changed in the predicted direction, four showed no change, and 15 changed in the opposite direction. Experimental *S*s as a group modified their answers in the predicted direction for 48 of the 65 adjectives, showed no change for five adjectives, and modified their answers in the non-predicted direction for 12 adjectives. Control *S*s as a group changed their answers in the predicted direction for 25 adjectives, showed no change for eight adjectives, and modified answers in the opposite direction for 32 adjectives.

2. *Acceptance of Fake Evaluation*

The response changes themselves indicate some degree of acceptance of the fake personality evaluation and there was no evidence to indicate that *S*s were suspicious or had discovered the purpose of the study. When the evaluations were given out the observers reported that there was some "uneasy" laughter, but *S*s did not attempt to show their evaluations to each other or discuss them. As the second inventory was turned in *S*s indicated on a scale of one to seven (1 = very low, 4 = average, 7 = very high) their degree of agreement with the personality evaluations. The median was 4.67, with agreement scores ranging from two to seven. This median is rather high in view of the extreme and arbitrary evaluations which were given the experimental *S*s. There was no relationship between the expressed degree of agreement and the amount of response modification.

3. *Positive Versus Negative Evaluations*

The effects of the two kinds of evaluations were tested by comparing response changes on those adjectives on which it was predicted that experimental *S*s would rate themselves more favorably with those on which they would rate themselves more unfavorably. For example, it was predicted that *S*s would rate themselves more favorably on adjectives related to lack of prejudice after being told they were open-minded and tolerant, and more unfavorably on items related to anxiety after being told they were extremely anxious. There was no significant difference between the responses to the two categories of adjectives ($t = 1.67$, $p > .10$), although the difference was in the direction of greater effectiveness for the negative than for the positive evaluations. This finding does not mean that negative evaluations are necessarily as effective as positive evaluations, since the personality dimensions being evaluated may differ widely in importance to the individual, but it does indicate that *S*s do not merely accept the pleasant evaluations of themselves and reject or ignore the unpleasant ones.

D. DISCUSSION

The results indicate that *S*s will accept fake personality evaluations from a supposedly authoritative source and will revise their self descriptions in accordance with them even though the evaluations reflect negatively on the *S*s. This acceptance does not appear to be a function only of the ambiguity of the terms used for evaluation since in the present study an attempt was made to be more specific than in the earlier studies by Forer (1) and Sundberg (3). Perhaps the critical ambiguity lies more in *S*'s understanding of himself and his view of the way in which he compares with others than in the evaluations themselves. *S*'s unsureness of himself and desire for more certainty, plus the related high interest in personality assessments could account for his readiness to entertain an evaluation which comes from a supposedly qualified source.

A question arises as to the factors involved in the revised self descriptions and several possibilities suggest themselves. Sarbin and Rosenberg (2) have described the adjective check list as a useful technique for obtaining a qualitative estimate of the self, and Trent et al (4) have reported on its merit for obtaining measurements of the self in a cross-cultural situation. One interpretation of the results of this study then is that the self concepts of the experimental *S*s actually did change somewhat in the light of the fake personality evaluation if one assumes that the self as seen by the individual is what is being measured. These college *S*s may also have been especially willing to entertain new hypotheses about themselves since some students at their college do appear to undergo a re-evaluation of their abilities and values.

Studies of the self have been criticized, however, on the grounds that the concept is a vague one which does not lend itself well to definition and measurement (5). Certainly if such a concept is open to modification in the short time interval and under the type of pressure used in the present study, a question is raised about the stability of the self structure. Unfortunately, it was not possible to gather data to assess the permanence of the induced change in self description since even if *S*s had not been informed of the deception, they would probably have discovered through subsequent discussions with each other that all had received identical evaluations.

Rather than make a possibly unwarranted assumption about any basic changes in the self structure, another interpretation of the findings might be as follows: *S*s did not see or feel themselves to be any different after reading the evaluations, but as a result of the inability to compare themselves with others, may subsequently have given descriptions of themselves which were more in line with those of a supposedly authoritative source. An individual knows whether he is taller or heavier than others, but he is less sure about comparisons on more covert dimensions such as sociability, self-regard, and anxiety tendencies. He may say, "I know that I feel anxious, but according to these scales I feel much more anxious than others do."

A related interpretation is that the individual does not see or feel himself to be any different, but may conclude that he does not have the necessary training or insight to evaluate himself correctly. In this case he may say to himself, "I don't really feel anxious, but someone who knows more about this than I do says that I have been misjudging myself."

Additional research is needed to determine which of these interpretations is the most useful and even whether or not it is possible to distinguish operationally among them. In research that involves the use of self appraisal as a dependent variable, e.g., in evaluating pre- and post-therapy changes, it would be particularly important to know to what extent the individual sees himself differently than he did before or to what extent he is merely describing himself differently.

E. SUMMARY

The present study reports the effect of a fake personality evaluation on subsequent self description. Sixty-eight under-graduates completed a person-

ality inventory which included 150 adjectives to be answered on a four-point scale according to the degree to which each adjective characterized the individual. Half of the *S*s (experimental group) subsequently received identical evaluations supposedly based on the personality inventory. The evaluation consisted of ten scales—scores on eight of which described the individual in either an extremely positive or an extremely negative way. All *S*s then completed a second questionnaire, which included the same 150 adjectives, and response changes from the first to the second admission were scored for each *S*.

The experimental group modified its self descriptions significantly ($p < .001$) in the direction of the fake evaluations while the control group showed no such modification. Negative evaluations had as much effect on the revised descriptions as did the positive evaluations. Several interpretations of the response change of the experimental group are offered.

REFERENCES

1. FORER, B. R. The fallacy of personal validation: a classroom demonstration of gullibility. *J. Abnorm. & Soc. Psychol.*, 1949, **44**, 118-123.
2. SARBIN, T. R., & ROSENBERG, B. G. Contributions to role-taking theory: IV. A method for obtaining a qualitative estimate of the self. *J. Soc. Psychol.*, 1955, **42**, 71-81.
3. SUNDBERG, N. D. The acceptability of "fake" versus "bona fide" personality test interpretations. *J. Abnorm. & Soc. Psychol.*, 1955, **50**, 145-147.
4. TRENT, R. D., FERNANDEZ-MARINA, R., & MALDONADO-SIERRA, E. D. The cross-cultural application of the adjectival check list adjustment index: a preliminary report. *J. Soc. Psychol.*, 1960, **51**, 265-276.
5. WYLIE, R. C. The Self Concept. Lincoln: Univ. Nebraska Press, 1961.

Department of Psychology and Education
Swarthmore College
Swarthmore, Pennsylvania

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 465-468.

FOOD INTAKE AS A FUNCTION OF DURATION OF FOOD DEPRIVATION*[1]

Wake Forest College and University of Richmond

ROBERT H. DUFORT AND JOHN H. WRIGHT

A. INTRODUCTION

In recent discussions of the status of research on hunger, Miller (3, 4, 5, 6) has cautioned against the use of food intake as a measure of the effects of different durations of food deprivation. He presented supporting data showing that intake following deprivation periods of moderate length increases with length of the deprivation period up to 24-30 hours and then fails to show additional increases. In discussing these data, he suggested that amount eaten, as a measure of severity of deprivation, is limited by the volume of the stomach and the ability of the *S* to deal with food.

The present study was designed to provide additional data on the relationship between food intake and food deprivation. In particular, it reports food intake following deprivation periods that are more severe than those customarily used. Intake during a two-hour eating period was measured following deprivation periods ranging from 0 hours to 192 hours in length.

B. METHOD

The *S*s were 44 male albino rats of the Holtzman strain, 213 days old at the beginning of the base period of the experiment. They were housed in individual 7 inches × 10 inches × 7 inches wire-mesh living cages in a soundproof laboratory room. Mean room temperature was 77 degrees F. The *S*s were exposed to the natural day-night cycle through a window in the room. They were maintained on Purina lab chow pellets and tap water.

The experiment started with an eight-day base period (3/10/61—3/17/61). During this time seven 24-hour food intake measures were recorded. On the last day of the base period, the *S*s were ranked on the basis of their mean 24-hour food intake and assigned randomly, eleven at a time, to one

---

* Received in the Editorial Office on February 1, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This paper is part of a thesis submitted by the junior author to the Department of Psychology, University of Richmond, in partial fulfillment of the requirements for the M. A. degree. The research was directed, and this report prepared, by the senior author. Publication of this research was supported by a grant to the senior author from the Graduate Council of Wake Forest College.

of eleven durations of food deprivation: 0, 12, 24, 36, 48, 72, 96, 120, 144, 168, and 192 hours. In this way, four matched groups of eleven *S*s each were formed; within each matched group, one *S* was in each deprivation condition.

A nine-day experimental period began on the day following the last day of the base period; all *S*s remained on ad libitum food and water from the last base-period day to the first experimental-period day (3/17/61—3/18/61). A two-hour eating period for all *S*s was set from 1:00 P.M. to 3:00 P.M. on the ninth day of the experimental period (3/26/61). Starting on the first day of the experimental period, food was removed from the *S*s in each deprivation condition at the appropriate number of hours before the common eating period. All *S*s continued to have free access to water during their deprivation. The *S*s in the 0-hour deprivation condition were not deprived of food prior to the eating period. At the beginning of the eating period, all *S*s were given fresh water and a supply of food pellets well in excess of their anticipated intake. Since the eating period started at 1:00 P.M., it was necessary to enter the experimental room twice at 1:00 A.M. to remove food from the 36-hour and the 12-hour *S*s; the activities of the *E* at these times were guided by weak flashlight.

C. Results and Discussion

An analysis of variance was performed on mean daily food intake during the base period to determine whether the matching procedure had resulted in equality on this measure for the *S*s in the different deprivation conditions. The analysis showed that the hypothesis of equality could be retained ($F = 1.07$; $df = 10, 30$; $p > .05$).

Figure 1 presents intake during the eating period as a function of length of deprivation. The function displays an initial sharp increase to 24 hours, a more gradual increase to 72 hours, and an apparent decline for the longer deprivation periods. Analysis of variance for the intake data of Figure 1 yielded a significant effect of length of deprivation ($F = 9.49$; $df = 10, 30$; $p < .001$). To determine the precise nature of this effect, Duncan's test was used to make systematic comparisons between the means of the individual deprivation conditions. This analysis showed that the 0-hour and the 12-hour deprivation conditions are significantly different from each other and from all other deprivation conditions ($p < .05$). However, for the remaining nine conditions beginning with the 24-hour condition, no significant differences between means were found.

In general, these findings support the contention that food intake is

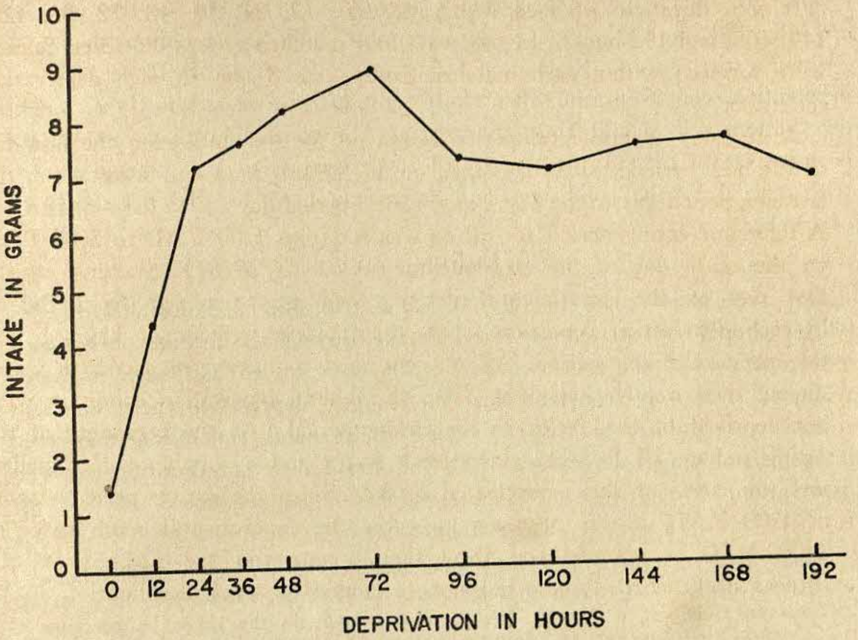


FIGURE 1
FOOD INTAKE DURING THE EATING PERIOD FOLLOWING DIFFERENT LENGTHS OF FOOD DEPRIVATION

not a suitable measure of the severity of food deprivation.[2] The failure of food intake to significantly increase for deprivations longer than 24 hours is in agreement with the results reported by Miller (3, 4) and by other investigators using the shorter deprivation periods (1, 2). This finding is also consistent with the operation of a stomach volume factor or some other factor that limits intake and prevents the *S* from reflecting in eating behavior the deficit produced by deprivation. While the present study provides no direct information about the nature of this limiting factor, the finding that there were no differences between the intakes of the severely deprived *S*s and the intake of the *S*s deprived for 24 hours suggests that the factor operates in a relatively constant way throughout the range of the longer deprivations.

Recently, Williams and Campbell (7) have measured intake following several of the longer deprivation intervals. They used deprivations of 24, 48, 144, and 192 hours and recorded milk consumption during a 24-hour eating period. They found that intake increased up to the 48-hour deprivation condition and then failed to increase more for the longer deprivations. It would thus appear that the conclusion that food intake is not a useful measure of severity of food deprivation has some generality with regard to the conditions and substances used for the determination of intake.

---

[2] The function in Figure 1 is apparently quite stable under the conditions of this experiment. A replication of the present study, using 22 *S*s in two matched groups of 11 *S*s each, produced a function highly similar in appearance.

D. SUMMARY

Rats were deprived of food for 0, 12, 24, 36, 48, 72, 96, 120, 144, 168, or 192 hours and tested for food intake during a subsequent two-hour eating period. Intake increased for deprivations up to 24 hours and then remained relatively constant for the longer deprivation periods. The results were interpreted as supporting the general conclusion that food intake is not a useful measure of the effects of different durations of food deprivation.

REFERENCES

1. BARE, J. K., & CICALA, G. Deprivation and time of testing as determinants of food intake. *J. Comp. & Physiol. Psychol.,* 1960, **53**, 151-154.
2. LAWRENCE, D. H., & MASON, W. A. Food intake in the rat as a function of deprivation intervals and feeding rhythms. *J. Comp. & Physiol. Psychol.,* 1955 **48**, 267-271.
3. MILLER, N. E. Shortcomings of food consumption as a measure of hunger; results from other behavioral techniques. *Ann N.Y. Acad. Sci.,* 1955, **63**, 141-143.
4. ———. Effects of drugs on motivation: The value of using a variety of measures. *Ann N.Y. Acad. Sci.,* 1956, **65**, 318-333.
5. ———. Experiments on motivation. *Science,* 1957, **126**, 1271-1278. (a)
6. ———. Objective techniques for studying motivational effects of drugs on animals. *In* S. Garattini and V. Ghetti (*Eds.*), *Psychotropic Drugs.* New York: Elsevier, 1957, pp. 83-103. (b)
7. WILLIAMS, R., & CAMPBELL, B. A. Weight loss and quinine-milk ingestion as measures of "hunger" in infant and adult rats. *J. Comp. & Physiol. Psychol.,* 1961, **54**, 220-222.

Department of Psychology
Wake Forest College
Winston-Salem, North Carolina

Department of Psychology
University of Virginia
Charlottesville, Virginia

Published as a separate and in *The Journal of Psychology*, 1962, **53**, 469-474.

VERBAL CONCEPT ATTAINMENT: A FUNCTION OF THE NUMBER AND STRENGTH OF POSITIVE INSTANCES PRESENTED*

University of Southern California and New York University

M. S. MAYZNER AND M. E. TRESSELT

A. PURPOSE

In verbal concept attainment *S* typically is confronted with an array of instances or words. Usually, some sub-set of the total array is made up of those instances that *E* has designated as positive concept instances (i.e., words that belong to the concept), while the remaining sub-set is made up of those instances that *E* has designated as negative concept instances (i.e., words that do not belong to the concept). The *S*'s task is to identify or separate the positive from the negative instances, and to respond to the positive instances with the correct concept name.

If one examines the preceding process in terms of an S-R mediational model first developed by Kendler and his associates (1, 2, 3) in this connection, each instance or word may be said to evoke implicit responses. When more than one instance evokes the same implicit response, the probability of *S* grouping these different instances together increases, resulting in *S* explicitly responding to these different instances with the same concept name. If the instances selected and the associated concept name corresponds to that sub-set of the total array that *E* has designated as positive concept instances, then *S* is said to have attained the concept. In a previous study (4), predictions based on this model concerning the effects of variations in the number of positive and negative instances presented were confirmed, with the results showing that increased numbers of positive instances facilitated concept attainment, while increased numbers of negative instances produced, in general, an opposite effect.

The present study is concerned with extending this S-R mediational approach further, by examining variations in the response strength of positive instances alone to evoke the correct implicit response. Specifically, it is predicted that in a concept attainment task in which positive instances only are presented one at a time to *S*, the number of positive instances required before the

---

* Received in the Editorial Office on February 1, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

concept is attained is critically dependent on the response strength of each positive instance to evoke the correct implicit response. Thus, with response strength high, only a few positive instances should be required to evoke the same implicit response and so attain the concept. With response strength low, however, many instances may be required to evoke the same implicit response and if response strength is very low, the same implicit response may never be evoked and so the concept may never be attained.

B. METHOD

1. *Subjects*

One hundred and twenty *S*s from introductory psychology classes at Barnard College, Columbia University were employed in the experiment.

2. *Material*

The stimulus material used in this study was selected from a word-list prepared by Mayzner and Tresselt (5) which gives the response frequencies (i.e., a measure of response strength) for a selected group of words judged for inclusion in a number of concepts. In the earlier scaling study (5), *S*s were presented with a list of 300 words and were required to judge each word for inclusion in some six concepts, namely, round, small, white, hard, smelly, and long. Response frequencies or the percentage of *S*s judging each word as belonging to each concept are available then as a measure of the response strength for each word to evoke a particular concept.

Three concepts, round, long and hard, were chosen for the present study. Also, words were selected to be employed as positive instances for these three concepts. For each concept, round, long and hard, 11 words each were chosen to fall within one of six response frequency ranges, namely, 11-25, 26-40, 41-55, 56-70, 71-85, and 86-100. The limit of 11 positive instances was fixed by the scaling study, in that for some of the response frequency ranges, only 11 words could be found within the prescribed range. For example, with the concept round, typical instances employed in the 11-25 range were—belt, cloud, fishhook; in the 26-40 range—cheese, eggshell, lips; in the 41-55 range—ellipse, thimble, waist; in the 56-70 range—bracelet, jar, pill; in the 71-85 range—clock, derby, equator; and in the 86-100 range—apple, doughnut, moon. It was assumed, that as response frequency ranges increase in value from 11-25 to 86-100, the response strength of the instances falling within that range to evoke a particular concept, also increases.

3. *Procedure*

There were six main experimental groups, one for each of the six different response frequency ranges, with 12 different *S*s per group. Each group received all three concepts, round (R), long (L), and hard (H), balanced for all different orders of presentation, i.e., two *S*s received RLH, two *S*s received RHL, two *S*s received LRH, etc. The *S*s in each group were told they were going to be shown a series of cards one at a time with words on them; all of these words would share a common sensory impression and could therefore be considered instances of the same category or concept, which they were to discover. They were then shown two simple examples. The first problem was then presented, in which the positive instances were exposed one at a tme, with those already presented left face up in order before *S*. *S* was instructed to respond, if he could, to each instance just once. Instances were given at the rate of one per minute until *S* gave the correct concept name, i.e., round, long or hard, or until all 11 instances had been presented, at which time *S* was given the correct solution. A different random order of the 11 instances was employed for all *S*s, with the one constraint, that each of the 11 different instances of each problem appeared at least once as the first instance presented to *S*.

Four additional groups of 12 different *S*s each also were employed, using the four lowest response frequency ranges, i.e., 11-25, 26-40, 41-55, and 56-70. These groups were run in an identical manner to the six main experimental groups with one exception, namely, *S*s were instructed to give as many different responses as they wished (constrained only by the one minute presentation rate) to each instance until the correct response was produced.

All *S*s were assigned randomly to the 10 different experimental groups, i.e., the six main experimental groups and the four additional groups. All responses made by *S* were recorded by *E* and the number of instances required by *S* before the correct concept name occurred was selected as the dependent response measure. If *S* did not produce the correct response with all 11 instances presented, he was given a score of 12.

C. RESULTS AND DISCUSSION

Since each *S* received three problems, to reduce any biasing of the results because of a particular problem, an examination was made first of the performance to the three concept problems. No systematic or significant differences were found and therefore each *S*'s median score for the three problems was selected for further analysis. The means, medians, standard deviations, and ranges of these median criterion scores are presented in Table 1 for the several experimental conditions of the study.

TABLE 1
CONCEPT ATTAINMENT SCORES FOR DIFFERENT RESPONSE FREQUENCY RANGES FOR SINGLE AND REPEATED RESPONSE CONDITIONS

	Response frequency ranges					
Single response	86-100	71-85	56-70	41-55	26-40	11-25
M	1.58	2.92	3.25	8.42	11.00	12.00
Mdn	1.50	2.00	3.00	12.00	12.00	12.00
SD	.64	1.66	1.53	4.50	2.04	0.00
R	1-3	1-6	1-7	2-12	5-12	12.12
Repeated response						
M			1.33	4.08	7.25	7.42
Mdn			1.00	3.00	6.00	10.00
SD			.59	2.63	3.51	4.82
R			1-3	1-10	2-12	1-12

The results in Table 1 show clearly the systematic increase in the number of positive instances required to attain the correct concept as response strength of the positive instances to evoke the correct concept decreases. For the single response condition, the increase is from a mean of 1.58 positive instances for the 86-100 response frequency range to a mean of 12.00 positive instances for the 11-25 response frequency range. An application of the Kruskal-Wallis Analysis of Variance by Ranks Test yields an $H = 47$, $df = 5$, and $P < .001$. Nonparametric techniques are used throughout the study because of the marked heterogeneity of variance among the several experimental conditions. For the repeated response condition similar results obtain, with an increase from a mean of 1.33 positive instances for the 56-70 response frequency range to a mean of 7.42 positive instances for the 11-25 response frequency range. An application of the Kruskal-Wallis Test yields an $H = 20$, $df = 3$, and $P < .001$. Also, as shown in Table 1, the number of positive instances required to attain the correct concept is less for the repeated response condition than for the single response condition for all four of the response frequency ranges examined (i.e., 1.33 vs. 3.25, 4.08 vs. 8.42, 7.25 vs. 11.00, and 7.42 vs. 12.00) and an application of the Mann-Whitney U-test to each of these four comparisons yield P-values well below the .05 level.

The results obtained from the single response condition support the main hypothesis of the study in showing that with high response strength of the positive instances the concept is attained with very few instances, while with low response strength, many instances are required or the concept may not be attained at all or at least not attained with the 11 instances which were available from the scaling study. A relevant study by Underwood and Richardson (6) employing a very different type of experiment design, in which *S*s were presented with four positive instances from each of six concepts, with the positive instances for two of the concepts possessing high response strength, two medium response strength and two low response strength, also has found a significant relationship between response strength and speed of concept attainment. These findings are quite compatible with the S-R mediational model of this study and while not treated by Underwood in these terms, may easily be done so. For example, in the Underwood and Richardson study, the four positive instances with high response strength rapidly evoke the same implicit response, thus yielding the correct concept name, while the four positive instances with medium or low response strength evoke the same implicit response only after many other stronger implicit responses have first been evoked, thus resulting in much slower attainment of the correct concept name.

Further evidence somewhat suggestive of the importance of implicit mediational responses may be found in the results of the repeated response condition. Here, as previously noted, significantly fewer positive instances are required to attain the correct concept name for each of the response frequency ranges examined. In the repeated response condition, since *S* is permitted to give as many explicit responses as he wishes during the one minute period, it seems plausible to assume that many more implicit responses are generated than in the single response condition. With the greater generation of implicit responses it is more probable then that the same implicit response will occur earlier, that is, yield the correct concept name with fewer positive instances.

D. SUMMARY

The present study has examined the effect of the number and strength of positive instances presented on verbal concept attainment. The results, predicted on the basis of an S-R mediational model, showed a significant and systematic increase in the number of positive instances required to attain the correct concept as the response strength of the positive instances to evoke the correct concept decreased.

REFERENCES

1. KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts in human concept-formation behavior. *J. Exper. Psychol.*, 1955, **49**, 165-174.
2. KENDLER, H. H., & KARASIK, A. D. Concept formation as a function of competition between response produced cues. *J. Exper. Psychol.*, 1958, **55**, 278-283.

3. KENDLER, H. H., & MAYZNER, M. S. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. Exper. Psychol.,* 1956, **51**, 244-248.

4. MAYZNER, M. S. Verbal concept attainment: A function of the number of positive and negative instances presented. *J. Exper. Psychol.* (in press).

5. MAYZNER, M. S., & TRESSELT, M. E. A comparison of judgmental and associational techniques in developing verbal concept formation materials. *J. Psychol.,* 1961, **51**, 331-342.

6. UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. Exper. Psychol.,* 1956, **51**, 229-238.

Department of Psychology
University of Southern California
Los Angeles 7, California

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 475-489.

THE EFFECT OF CONVERGENCE ON PERCEIVED SIZE AND DISTANCE*[1]

U.S. Army Medical Research Laboratory, Fort Knox, Kentucky

WALTER C. GOGEL

A. INTRODUCTION

The problem of the effect of the convergence of the eyes upon perceived size and distance has received considerable experimental attention. Summaries of research on this problem are found, for example, in Linksz (10), Osgood (12), and in Woodworth and Schlosberg (14). However, in this research, it is not always clear whether absolute or relative perceived size and distance are being measured, and whether absolute convergence or convergence change is the stimulus variable being studied. The distinctions between absolute and relative stimuli and between absolute and relative perceptions are important if appropriate conclusions are to be made from the experimental data. Perceived absolute distance is the *O's* perception that an object is located at a particular distance from himself. A perception of this type is sometimes described as involving egocentric localization. This is to be contrasted with a perception of relative distance which is the perceived distance between objects. Similarly the perception of absolute size is the perception of the absolute extent of an object. This is to be contrasted with a perception of relative size which is the perceived ratio of one object size with respect to another. The absolute convergence of the eyes is the angle formed between the visual axes when the eyes are positioned to bring the image of a binocularly fixated object on corresponding points on the fovea of each eye. This is to be contrasted with a convergence difference which is a change in this angle resulting from a change in the distance of the fixated object.

The present study will be concerned with the effect of absolute convergence on perceived absolute size and distance. Two recent studies have considered the relation between absolute convergence and perceived absolute distance (3, 4). To measure perceived absolute distance, a visual "ruler" was provided which consisted of a monocularly viewed alley against

---

* Received in the Editorial Office on February 9, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to thank Don L. Carmichael for his assistance in collecting and analyzing the data of this study.

which the perceived distance of a binocularly viewed object could be judged. It was found that only for a minority of *O*s was there any evidence that perceived absolute distance varied as a function of absolute convergence. Even with these *O*s, the discrimination of different distances was very limited. Another study (5) measured the possible relation between absolute convergence and perceived absolute size. The method of measuring perceived absolute size consisted of using a kinesthetic adjustment to duplicate the apparent size of a constant retinal size at different convergences. It was found that absolute convergence was not a sensitive determiner of perceived absolute size. A change in perceived absolute size as a consequence of a change in convergence, if it occurred, was limited. In each of these three studies the convergence values used were usually between 0 and 4 degrees. While this established that absolute convergence, if a cue at all, is relatively insensitive, it is possible that with greater values of absolute convergence, perceived absolute size and distance would be more frequently and more precisely determined. In the present study, the same methods of measuring perceived absolute size and distance were used but with convergence values from 0 to 12 degrees.

B. METHOD

1. *Apparatus*

a. The binocularly observed object. The binocularly observed object was the same as that used in a previous study (5). It consisted of a luminous horizontal line segmented by three short vertical lines (stakes). This object had an over-all angular width of 2°54′20″ and was presented frontally to *O* at each of five convergence values (0°, 3°, 6°, 9°, and 12°). The binocular object was stereoscopically generated by using two independent optical systems, one for each eye. By rotating these optical systems in opposite directions, the direction at which the light stimulus entered each eye could be controlled by *E*. This permitted *E* to adjust the stereoscopic object to the required convergence value. In addition to the binocular portions of the object, small monocular components were added so that, when *O* properly fused the binocular object, a unique configuration of the stimulus was perceived (see Figure 1 of Reference 5). In this way, *E* could be certain that the eyes of *O* were at the required convergence value. The optical systems were such that, regardless of the convergence value to the binocular object, the light coming from the binocular object was always parallel, i.e., always the same as that from a very distant object. The method used in generating a similar stereoscopic object is discussed in more detail elsewhere (2).

b. Apparatus for the measurement of perceived absolute size. The apparatus for measuring the perceived absolute size of the binocular object at each of the convergence values consisted of two aluminum rods located at about the level of *O*'s waist. The right rod was stationary while the left rod could be moved laterally. Each rod was ½ inch in diameter and approximately six inches high. A measuring stick attached to the movable rod permitted *E* to measure the lateral separation between the two rods. This apparatus was identical to that used in a previous experiment (5).

Throughout this study, the eye pieces through which *O* viewed the binocular object each contained a pinhole 0.5 mm. in diameter. This permitted *O* to clearly see the binocular object regardless of its convergence value. If the pinholes had not been present, the binocular object, by requiring the accommodation appropriate for a distant object, could have appeared blurred for the large convergence values. This could have resulted as a consequence of the dissociation of the normal accommodative-convergence relation. With the pinholes however this did not occur. The pinholes were placed as closely as possible to *O*'s eyes and no part of the binocular view to the object was obscured for any value of convergence.

c. Apparatus for the measurement of perceived absolute distance. The apparatus used here to measure perceived absolute distance was very similar to that used in previous studies (3, 4). It consisted of a monocularly viewed alley 34.2 feet long and 4.2 feet wide. Sixteen white numbered cards (also viewed monocularly) were distributed uniformly in distance along the alley, with the nearest card (Card no. 1) at 35 inches from *O* and with 26 inches between cards. A schematic perspective view of the alley is shown in Figure 1. This figure also contains a schematic representation of the binocular object (the three-stake reticle). When visible, the alley and numbered cards were always seen only through the pinhole aperture before the right eye. While *O* saw both the binocular object and the alley with his right eye, only the binocular object was visible with his left eye. Except for the binocular object, the left eye of *O* was always in complete darkness.

The monocular alley was visible only when a black curtain in front of the stereoscopic instrument was raised. A black velveteen hood was located at the *O*'s position as is illustrated in Figure 1. With the front curtain lowered, the light sources for the binocular object turned off, and the hood adjusted in place, *O*'s position was in complete darkness.

2. *Observers*

Twelve men were used as *O*s throughout this study. Each *O* had a stereoscopic acuity of at least 75 seconds of arc as measured with the Diagnostic Series of the Keystone Telebinocular.

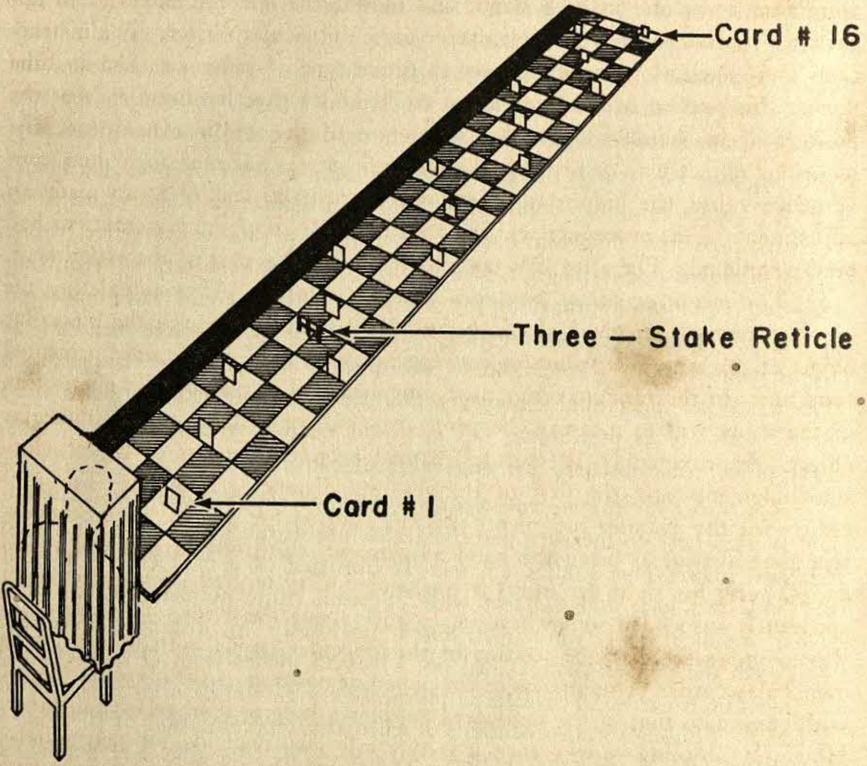


FIGURE 1
A Schematic Perspective Diagram of the Alley

3. *Procedure*

a. The measurement of perceived absolute size. The five values of convergence were each presented four times in random order to each *O*. The general procedure was as follows. Each *O* was carefully instructed as to his task with a small model (10 cm. wide) of the binocular object used to illustrate the shape of the object which would be perceived when the proper binocular fusion had occurred. The *O* was seated before the instrument and the proper convergence value was adjusted by *E* with the light sources for the binocular object turned off. Following this, the binocular object was turned on, *O* checked to see that the binocular object was visible with each eye, and reported when the proper binocular fusion had occurred. The *O* grasped the aluminum rods, one in each hand, and moved the left rod laterally to apparently reproduce the right-left extent of the binocular object. The instructions were such as to emphasize an objective type of judgment. At no time during this portion of the experiment could *O* see the aluminum rods or the position of his hands. Upon the completion of the width adjustments, the binocular object was turned off, the instrument was adjusted for a new convergence value, the binocular object was turned on, and *O* again made an adjustment. This procedure was repeated until all 20 of the presentations had been completed. The alley was not visible during this part of the experiment.

b. The measurement of perceived absolute distance. After completing the measurement of perceived absolute size, *O* was presented with the binocular object at the same five values of convergence with each convergence presented four times in the random order used previously. The purpose of this portion of the study was to measure the perceived absolute distance of the binocular object. Approximately 10 minutes elapsed between the last of the absolute size judgments and the first of the absolute distance judgments. The procedure for the distance judgments was the same as for the size judgments except that instead of using the hand adjustment, the front curtain was raised and *O* with his right eye saw the monocular alley with the binocular object apparently suspended somewhere in distance above the cards. The task of *O* was to judge the distance position of the binocular object with respect to the monocular cards. A judgment of the apparent position of the binocular object with respect to that of the numbered cards is a judgment of relative distance. However, absolute convergence was the only cue available for perceptually localizing the binocular object in the monocular alley. Therefore, the perception of the distance position of the binocular object relative to that of the numbered cards must be regarded as involving an absolute distance judgment. In other words, there were no relative cues between the binocular object and the monocular cards to relate the perceived distance of one to that of the other. Therefore, the only cues which could be used for this judgment were cues relating each to the position of the *O*. Thus, the use of the monocular alley was a valid method of measuring the perceived absolute distance of the binocular object.

In judging the apparent position of the binocular object with respect to the monocular cards, the *O* was permitted to use 1/4 units of the apparent dis-

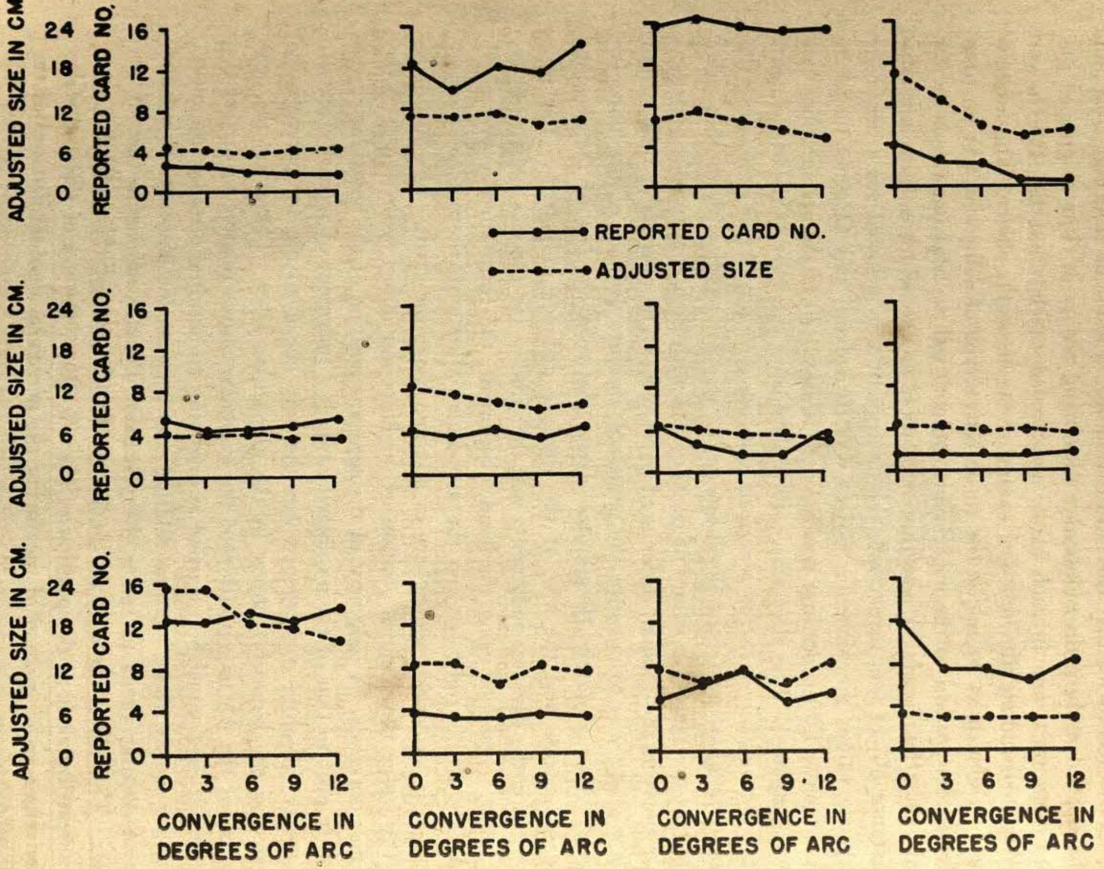


FIGURE 2
Perceived Size and Distance as a Function of Convergence for an Object of Constant Angular Size for Each *O*

tance between the successive cards and if necessary to extrapolate the scale of numbered cards beyond Card 16 and closer than Card 1. The 0.5 mm pinhole apertures were used here also. This resulted in both the binocular object and the entire length of the monocular alley being always in clear vision.

Following the completion of the perceived distance judgments, the *O*s were asked to estimate the distance in feet to each of the monocular numbered cards. The order in which the cards were designated for this task was randomly determined for each *O*. The purpose of these estimates was to be certain that the monocular alley appeared to *O* to extend away from himself in distance.

C. RESULTS

1. *Results from the Absolute Size Judgments*

The average results from the kinesthetic size adjustments for each *O* as a function of convergence are shown by the dashed line curves of Figure 2. Since the binocular object was of constant angular size it should have appeared larger with the smaller values of convergence (greater distances). Thus, in Figure 2, a relation between absolute convergence and perceived absolute size would be indicated by the dashed lines sloping downward to the right. It will be noted that only about half of the *O*s showed a relation of this type and, even for these, the change in perceived size was small compared with the change in physical size calculated to produce the constant angular size at the different convergence distances. This calculated or theoretical curve is shown by the dashed line curve of Figure 3. The average of the twelve "adjusted size" curves of Figure 2 is given by the curve labeled "obtained size" in Figure 3. When applied to the dimension of perceived size, the ordinate of Figure 3 ("Extent") gives the value of the kinesthetic size adjustment in centimeters.

Table 1 shows the results from an analysis of variance of the kinesthetic size adjustments using the results from all *O*s. It appears that the changes in the average adjusted size as a function of convergence, though small, were nevertheless statistically significant.

TABLE 1
SUMMARY OF ANALYSIS OF VARIANCE OF "KINESTHETIC" DATA

Source	df	Mean square	F
Convergence	4	13.60	6.65**
Observers	11	86.96	42.52**
Error	44	2.05	
Total	59		

** Significant at the .01 level

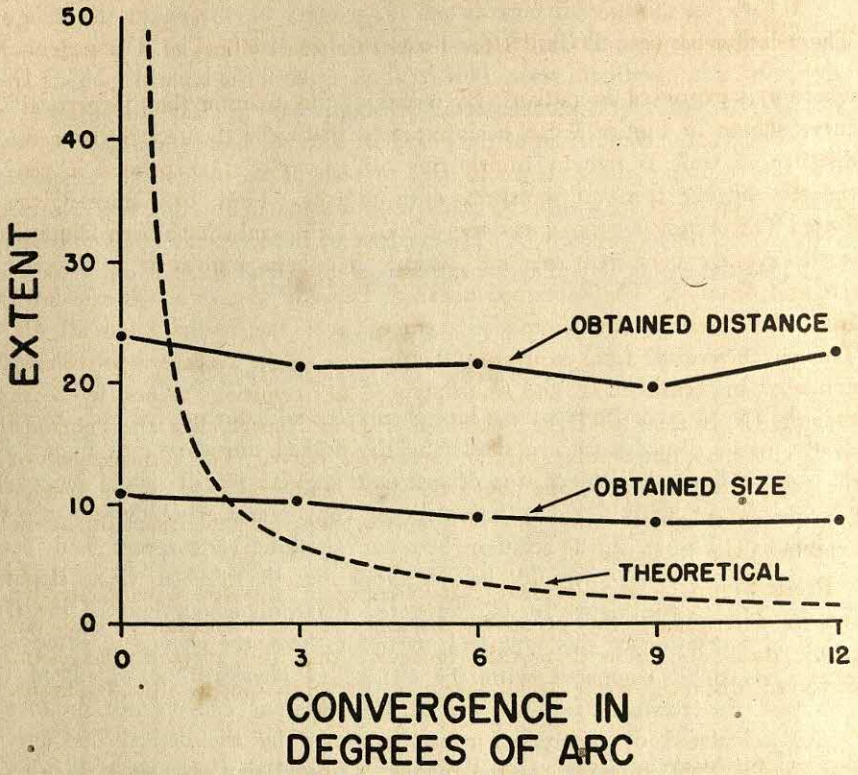


FIGURE 3
PERCEIVED AND PHYSICAL SIZE AND DISTANCE AS A FUNCTION OF CONVERGENCE FOR AN OBJECT OF CONSTANT ANGULAR SIZE

2. *Results from the Absolute Distance Judgments*

The average results from the distance judgments for each *O* as a function of the convergence of the eyes are shown by the solid line curves of Figure 2. The binocular object should have appeared to be at a greater distance for smaller values of convergence. The greater the reported card number the larger was the perceived distance. Thus, in Figure 2, a relation between absolute convergence and perceived absolute distance would be indicated by the solid line curves sloping downward to the right. Only four or five of the *O*s showed any relation of this type and, even for these *O*s, the change in perceived distance was usually small.

For an object of constant angular size (θ) both its physical size (S) and its physical distance (D) are inversely related to absolute convergence.

The relation between D and S for reasonably small values of θ is $\theta = \frac{S}{D}$, where θ is expressed in radians. By means of this relation the "theoretical" curve shown in Figure 3 can be interpreted to be the theoretical curve for distance as well as size by multiplying all values of D expressed in centimeters by the constant θ expressed in radians. Thus, in Figure 3, the dotted (theoretical) curve gives the values which would have been obtained if the average perceived size and distance had been equivalent to physical size and distance. The "obtained distance" curve of Figure 3 was calculated from the reported card numbers averaged over the results from all *O*s. This curve resulted from expressing the distance to the average reported card numbers in centimeters and multiplying the resulting values by θ in radians. It is clear that the "obtained distance" curve, like the "obtained size" curve, is not similar to the theoretical curve. The displacement of the two obtained curves, one above the other, indicates that on the average the binocular object was perceived to be more distant than might be expected from its perceived size.

From Figure 3, it is evident that convergence was not a very effective cue in determining the perceived distance of the binocular object. An examination of Table 2 reveals, however, that the average changes in perceived distance, while small, were nevertheless statistically significant.

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE OF "REPORTED CARD" DATA

Source	df	Mean square	F
Convergence	4	2.95	3.01*
Observers	11	113.26	115.37**
Error	44	.98	
Total	59		

* Significant at the .05 level
** Significant at the .01 level

The median results from the estimates in feet to each of the monocularly viewed cards is shown in Figure 4. Medians rather than means are used since the data tend to be skewed. It is apparent that the estimated distance to the cards in feet is in general linearly related to the number (distance) of the cards. It is evident that the *O*s perceived the monocular alley as extending away from themselves in depth and that, therefore, the alley constituted an adequate ruler for measuring the perceived absolute distance of the binocular object.

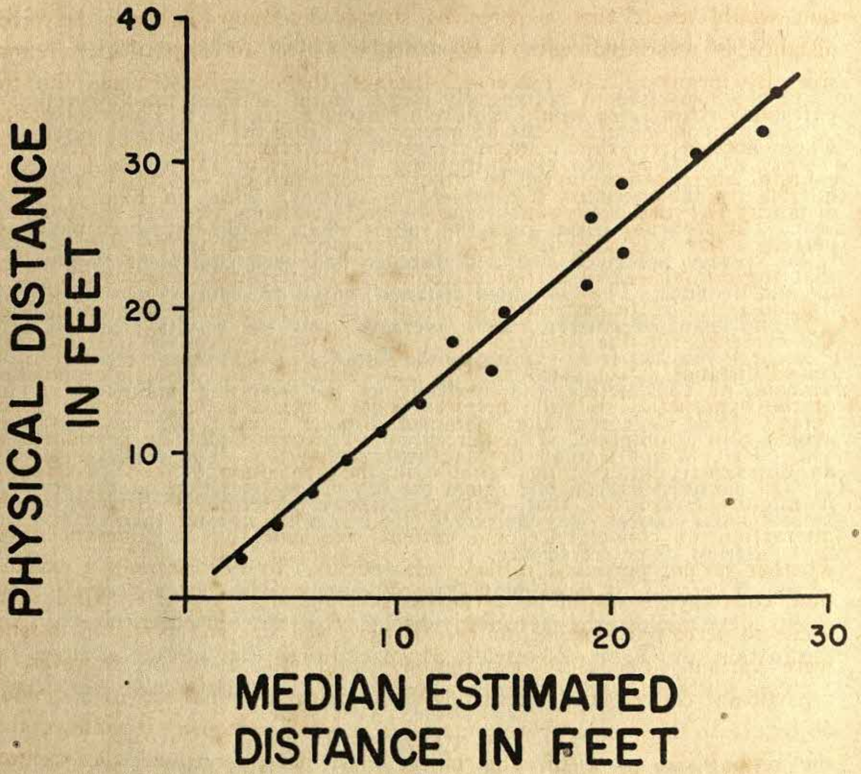


FIGURE 4
MEDIAN ESTIMATED DISTANCE AS A FUNCTION OF PHYSICAL DISTANCE IN THE MONOCULARLY VIEWED ALLEY

D. DISCUSSION

The reason for expecting a relation between convergence and perceived distance is clear. The convergence of the eyes is the angle formed between the visual axes when the eyes are positioned so as to bring the image of a binocularly fixated object on corresponding points on the fovea of each eye. Thus, proprioceptive impulses from receptors in the eye muscles producing this positioning (see 8) could be a cue to the distance of the fixated object. The reason for expecting a relation between convergence and perceived size is less clear. Obviously, there is no relation between the physical size of an object and its convergence distance from the eyes. However, the possible relation between convergence and perceived size might be thought to come about in one of two ways. One possible explanation would assert that a perceived size is determined when a perceived distance is associated with the particular object of a particular retinal size. By means of this perceived distance, the object associated with the particular retinal size would acquire a perceived size (13). This explanation when applied to the relation between convergence and perceived size contains three assumptions: (*a*) that convergence can determine perceived distance; (*b*) that a necessary relation exists between perceived distance and perceived size, e.g., the size-distance invariance hypothesis (see 9); and (*c*) that the retinal size of a stimulus object has perceptual consequences.

Evidence for the second assumption contained in this explanation, e.g., the evidence for the assumed invariance between perceived size and perceived distance, is equivocal (1, 9). However, there is a possible explanation of the hypothetical relation between convergence and perceived size which avoids this assumption. The perception of distance may be considered as an unnecessary intervening variable in the perception of size. Specifically, it might be asserted that perceived size is determined directly by the interaction of convergence and retinal size and thus is independent of whether or not perceived distance also occurs. In this alternative explanation, convergence would be a cue to perceived size even if it failed to be a cue to perceived distance, or even if perceived size and perceived distance were variant. The third assumption would still be made, however. If a particular convergence together with a particular retinal size of a stimulus object is to be sufficient to determine a perceived size, it follows that the retinal size of a stimulus object must have perceptual consequences. It seems therefore that the hypothesis that absolute convergence determines perceived size, regardless of whether it determines perceived distance, contains the assumption that the absolute size of a retinal image has perceptual consequences. However, as has been discussed in a previous study (5), there is some evidence against the assumption of a relation between absolute retinal size and perception. Therefore, the general inadequacy of convergence as a determiner of perceived size which was experimentally found in the present study is not completely unexpected.

Using a convergence range of 0 to 12 degrees in the present study instead of the 0 to 4 degrees used previously (3, 4, 5) did not seem to increase substantially the proportion of *O*s who evidenced a relation between convergence and either perceived absolute size or distance. It seems that absolute convergence is not a very adequate cue to perceived absolute size or perceived absolute distance. This means that a perception of relative size as a function, for example, of convergence differences should not be explained as a difference between two perceptions of absolute size from convergence per se. It also means that a perception of relative depth in binocular situations should not be explained as a difference between two perceptions of absolute distance from convergence per se.

If the same *O*s who showed distance discriminations in Figure 2 had also shown size discriminations, this would be evidence for the validity of the size-distance invariance hypothesis. It will be seen that this is only true for the results given in the last figure of the upper row of Figure 2. The first figure in Row 3, for example, shows size but not distance discriminations. However, since the size and distance of the binocular object were so poorly discriminated, this study clearly is not an adequate test of the size-distance invariance hypothesis.

The size adjustments from the present study and from the previous one (5) differ in one general aspect. The over-all average size adjustment was 24.8 cm. in the previous study and is 9.5 cm. in the present study. Since convergence did not appear to be a very effective cue, this difference cannot be attributed to the greater convergence values used in the present study. Perhaps this difference is attributable to the use of the pinhole apertures in the present but not the previous study. It has been pointed out (7, p. 128; 11) that with pinhole observation the size of the retinal image is reduced when the eye is accommodated for a point nearer than the object being viewed. In the present study, the accommodation to the binocular object was equivalent to that for a very distant object. With increasing convergence the accommodation of the eyes could possibly have increased [convergent-accommodation (6)] with the result that the retinal size of the binocular object might have become smaller with the greater convergences. This change in retinal size, if it occurred, would be expected to be small (see 6). It would be in the direction of producing a smaller perceived size for a greater convergence. Thus, this effect might have reduced somewhat the average perceived size. It might also be considered to be a possible explanation of why some *O*s showed some amount of size discrimination in the study. However, since the effect was in the direction of producing a decrease in perceived size with an increase in convergence, it cannot be used to explain why the size discriminations did not occur with greater frequency or sensitivity.

An attempt was made to measure the effect of pinhole vision on retinal size as a function of the accommodation of the eyes. For this purpose, as in Maier's study (11), the pinhole was used before only one eye of *O*. The three-stake reticle was seen only through the pinhole and the visual alley was presented binocularly. A binocularly viewed screen containing adequate cues to accommodation was placed successively at the distances required for approximately 12, 9, 6, 3, and 0 degrees of convergence. The retinal size of the three-stake reticle in the right eye was measured by *O* indicating its size on the screen when the screen was at each of the distances. Under these conditions, both the accommodation of the left eye and the convergence to the screen would operate to change the accommodation of the eye viewing through the pinhole. The average results of these measurements for eight *O*s was that the retinal size was 2° 56′, 2° 58′, 2° 59′, 2° 56′, and 2° 58′ at the approximately 12, 9, 6, 3, and 0 degrees of convergence respectively. It appears that the effect of pinhole vision on retinal size, if it occurred in this study, was very small. One reason for the smallness of this effect was that the pinholes were located very close to the *O*'s eyes throughout the present study.

The average adjusted size in the present experiment and the width of the small model used in the size instructions are very similar. This must be regarded as fortuitous. The same model and same instructions were used for the size adjustments in the present study and in the study which gave the 24.8 cm. average size adjustment (5). In both studies, the instructions were such as to emphasize that the size of the model had no necessary relationship to the size of the stimulus objects.

The *O*s whose results are shown in the 1st and 4th columns of Row 1 of Figure 2 are identified as *O*s A and D in previous studies (3, 4, 5). Both here and in the two previous studies involving perceived distance (3, 4), these *O*s saw the binocular object as being at a greater distance with a smaller convergence. In Figure 2, each decrease in convergence for these two *O*s resulted in an average increase in perceived distance. Evidently, the elimination of the cues of accommodation in the monocularly viewed alley by using the pinholes did not eliminate the distance discriminations. This provides further evidence that, with a few (experienced) *O*s, absolute convergence can be a cue, though an insensitive cue, to perceived absolute distance.

In a previous study (5) it was argued that the small discriminations of perceived size which occurred as a function of convergence were a consequence of the successive presentations, i.e., they were relative judgments made over a period of time. In the present study the time between turning off one convergence stimulus and presenting another was approximately 1/2 minute. It is, therefore, possible that the same explanation of the size discriminations in the present study could be applied. However, this is not the case with the distance discriminations since a previous study has shown that a twenty-minute interval between presentations did not reduce the ability to make such discriminations (3). Whether the size discriminations which were present would also remain when larger time intervals occurred between presentations is not known.

E. Summary and Conclusions

The role of absolute convergence in the perception of absolute size and distance was investigated using convergence values from 0 to 12 degrees. Perceived absolute size was measured by means of a kinesthetic adjustment. With this method, the apparent width of a binocular object of constant angular size was measured at the different convergence values. Perceived absolute distance was measured by providing a visual ruler consisting of a monocularly observed alley containing numbered rectangles. The *O* judged the apparent position of the binocular object with respect to the rectangles for the values of convergence previously used with the size judgments. The results indicate that convergence was a significant contributor to perceived size and distance for some *O*s. However, even with the *O*s who showed some effect of convergence, the sensitivity in making the size and distance discriminations was low. In general, the study supports the conclusion that, even in situations in which no conflicting cues are present, absolute convergence is an imprecise and usually negligible determiner of perceived size and distance.

References

1. Epstein, W., Park, J., & Casey, A. The current status of the size-distance hypothesis. *Psychol. Bull.* 1961, **58**, 491-514.
2. Gogel, W. C. An observer constant in the perception of stereoscopic depth. USAMRL Report No. 316, 1958, 1-24, Fort Knox, Kentucky.
3. ———. Convergence as a cue to absolute distance. *J. Psychol.,* 1961, **52**, 287-301.
4. ———. Convergence as a cue to the perceived distance of objects in a binocular configuration. *J. Psychol.,* 1961, **52**, 303-315.
5. ———. Convergence as a determiner of perceived absolute size. *J. Psychol.,* 1962, **53**, 91-104.
6. Heinemann, E. G., Tulving, E., & Nachmias, J. The effect of oculomotor adjustments on apparent size. *Amer. J. Psychol.,* 1959, **72**, 32-45.
7. Helmholtz, H. Treatise on Physiological Optics. Vol. I (J. P. C. Southall, Ed.) Manasha, Wisconsin: Banta, 1924, 127-128.
8. Irvine, R. S. New evidence relating to proprioceptive sense in the extrinsic eye muscles. *Amer. J. Ophthal.,* 1950, Series 3.33, 479-481.
9. Kilpatrick, F. P., & Ittelson, W. H. The size-distance invariance hypothesis. *Psychol. Rev.,* 1953, **60**, 223-231.

10. LINKSZ, A. Physiology of the Eye. Vol. 2, Vision, New York: Grune & Stratton, 1952.
11. MAIER, N. R. F. The illusion of size in pin-hole vision. *Amer. J. Psychol.*, 1929, **41**, 291-295.
12. OSGOOD, C. E. Method and Theory in Experimental Psychology. New York: Oxford Univ. Press, 1953.
13. SCHLOSBERG, H. A note on depth perception, size constancy, and related topics. *Psychol. Rev.*, 1950, **57**, 314-317.
14. WOODWORTH, R. S., & SCHLOSBERG, H. Experimental Psychology. New York: Holt, 1954.

Civil Aeromedical Research Institute
Federal Aviation Agency
Aeronautical Center
P.O. Box 1082
Oklahoma City, Oklahoma

Published as a separate and in *The Journal of Psychology,* 1962, **53**, 491-497.

THE EFFECT OF SYSTEMATIC AND NONSYSTEMATIC PRESENTATION OF STIMULI ON PERIPHERAL VISUAL ACUITY*[1]

Department of Psychology, Carnegie Institute of Technology

HARRY W. KARN, LEE W. GREGG, AND GORDON F. PITZ

A. INTRODUCTION

The present investigation is one of a series of studies concerned with the role of experiential variables on the perceptual response. The basic methodology involves a determination of certain measures of visual performance under conditions characterized by the introduction of a variety of specific cues capable of being learned (2).

Studies of visual acuity typically fall within the area of the classical pure sensory experiment where the emphasis has been upon the manipulation of physical stimulus variables and the role of experiential variables delimited by restriction of *S*'s responses through instruction and preliminary practice. However, the fact remains that a variety of momentary influences including attention factors and the categorical judgments required of *S* may be determined by prior experience. Thus, the acuity experiment is in fact a study in perception.

Viewing the acuity response as a perceptual phenomenon, the present experiment was designed to determine the effect of a specific perceptual cue on acuity measures in the peripheral retina. The critical feature of the experiment was the systematic and nonsystematic presentation of acuity targets at varying exposure times. Differences obtained on systematic vs. nonsystematic modes of presentation would indicate the presence of a specific experiential factor contributing to the perceptual act.

B. METHOD

1. *Subjects*

Thirty-two college students with normal vision (without glasses), eight in each of four treatment groups, served as *S*s.

---

* Received in the Editorial Office on February 14, 1962, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This research was supported by the Air Force Office of Scientific Research of the Office of Aerospace Research under Contract No. AF 49(638)-770.

2. *Apparatus*

The stimuli in the form of Landolt rings were projected onto a hemispherical screen 10 feet in diameter constructed so that all surface points were equidistant from the eye of the observer. The surface of the screen was finished with a flat white paint and a 1/4 of an inch diameter fixation spot was recessed into the screen at the central point of regard. The screen was uniformly illuminated to provide a homogeneous field with an illumination level of 38.58 mls. A panel covered the open face of the hemisphere except for openings for *S*'s headrest and for projection of the stimuli. The viewing distance from *S*'s eye to the screen was 5 feet.

A projector fitted with a shutter was used to project a series of stimulus slides consisting of Landolt rings with an outside diameter of .75 inches and an inside diameter of .47 inches and gap sizes of .035, .059, and .082 inches. The brightness of the Landolt rings was 241.14 mls.

3. *Procedure*

The Landolt rings were foveally and peripherally presented to all *S*s in blocks of 24, within which blocks, size and location of the gap varied randomly. The gap appeared each time in one of the four quadrants of the Landolt ring.

For all *S*s, the first block of stimuli was presented foveally at a duration of 1 second. The duration of all subseqent blocks of presentations varied for the four groups; 1 second for Group 1; .5 seconds for Group 2; .2 seconds for Group 3; .1 second for Group 4. The second block of stimuli was presented foveally, then four blocks were presented peripherally five degrees out from the fixation spot of the screen; the seventh block was presented foveally, and finally four more blocks were presented at five degrees. In all, 11 blocks of 24 stimuli each were presented to each *S*.

The stimuli that were presented peripherally were ordered in two ways: "systematically," in which all 24 presentations in a block appeared at the same point on the screen, either above the fixation spot, below the fixation spot, to the left or to the right of the fixation spot; "nonsystematically," in which the position of the stimulus in relation to the fixation spot, varied at random among the four points within each block.

For half of the *S*s in each group, systematic blocks were Blocks 3 through 6, and the nonsystematic blocks were Blocks 8 through 11. For the other half of the *S*s, this order was reversed. The four systematic blocks, the four positions of the target stimuli, and the four *S*s in any one group who were given the systematic blocks at the same time, formed a latin square.

All of the *S*s viewed the stimuli monocularly with the left eye. An opaque patch covered the right eye. With *S* seated in front of the hemisphere and his head in the headrest, he was told that a circle containing a gap would be flashed below the fixation spot, following a ready signal. He was to indicate where the gap appeared by pressing a button on a response panel. The *S*s were told to guess about the location of the gap if unsure.

Following the initial instructions, the first two blocks of foveal stimuli were presented. Before presenting the peripheral stimuli *S*s were told that the circle with the gap would be moved away from the center but that they should carefully maintain their fixation on the central spot and not attempt to look directly at the circle. Under these conditions, if foveal fixation was maintained, the circles fell five degrees out from the fovea. *S*s were told before each systematic block where the circle would appear, i.e., above, below, to the left or to the right of the fixation spot.

C. Results

The data were analyzed in terms of number of errors made by each *S* within each block of 24 stimulus presentations. Since there were four alternative responses to each stimulus, 18 errors per block represents chance performance. Blocks 1, 2 and 7 (concerned with foveal determinations) provide a check on the stability of acuity measurements in the fovea throughout the entire series of determinations. Errors made on the first block, where the stimuli were presented foveally for one second for all groups, were analyzed separately. There were no significant differences between any of the groups or subgroups. In the analysis of Blocks 2 and 7, where duration varied across groups, duration produced a significant effect ($F = 17.48$; $df = 3/24$), mean total errors decreasing as follows: 9.06, 4.69, 3.69, and 2.25 for the exposure times .1, .2, .5, and 1 second, respectively. No other effects were significant.

Of major interest is the way in which the data from the peripheral determinations comprising Blocks 3, 4, 5, 6, 8, 9, 10, and 11 vary as a function of duration and mode of presentation (systematic and nonsystematic). Figure 1 shows a differential error rate between systematically and nonsystematically presented targets at exposure times of .1 and .2 seconds, which disappears at the longer durations. Also shown in Figure 1 is a sharp decrease in foveal error scores from .1 to .2 seconds exposure and a gradual decline thereafter.

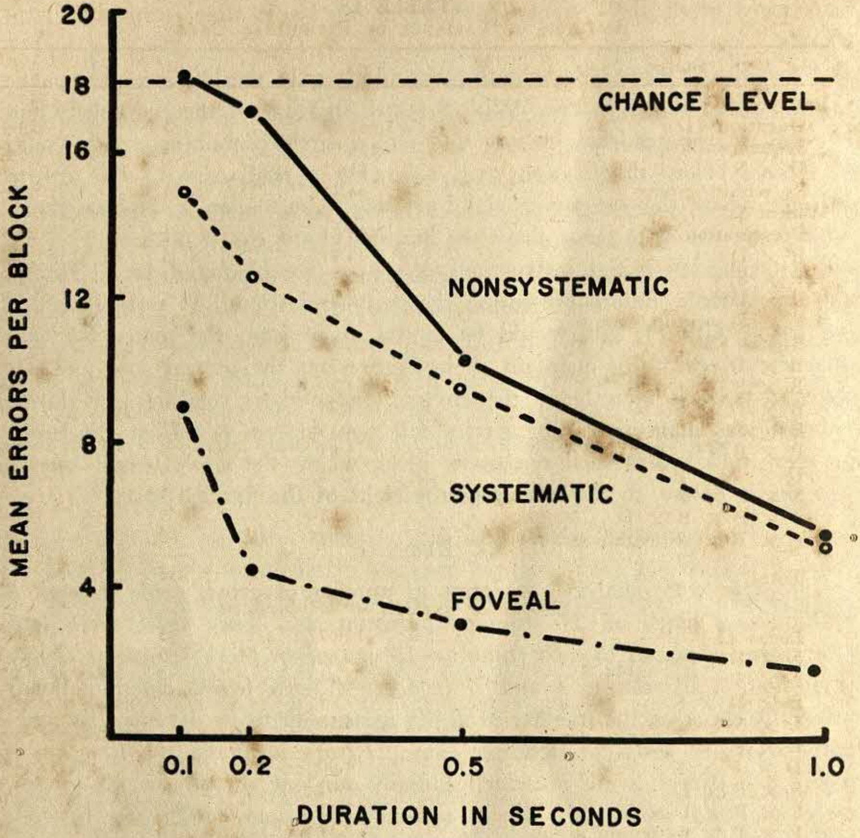


FIGURE 1
CHANGES IN ACUITY RESPONSE AS A FUNCTION OF DURATION AND MODE OF PRESENTATION

An analysis of variance of the peripheral data is summarized in Table 1. The sources of variance are as follows: (*a*) duration, the exposure time for stimuli which varied across the four groups of *S*s; (*b*) sequence, the order in which the systematic and nonsystematic blocks occurred within subgroups; (*c*) presentation, systematic or nonsystematic; (*d*) blocks, groups of 24 stimulus presentations; (*e*) locus, position of circles in the up, down, left or right quadrants.

The main effects of duration and presentation are significant as is the D × P interaction. Significant differences also occur among blocks. An alternative analysis of block effects indicated a significant locus effect and significant L × D, L × P and S × P × L interactions. These values

TABLE 1
Analysis of Variance of Peripheral Data

Source	*ms*	*df*	*F*
Between *S*s			
Duration (D)	1,641.37	3	24.44*
Sequence (S)	136.60	1	2.04
D × S	31.01	3	.46
*S*s within groups	66.85	24	—
Within *S*s			
Presentation (P)	304.04	1	24.56*
P × D	64.53	3	5.21*
P × S	12.71	1	1.03
P × D × S	12.04	3	.97
P × *S*s within groups	12.38	24	—
Blocks (B)	22.41	3	4.93*
B × D	6.55	9	1.44
B × S	1.39	3	.31
B × D × S	4.25	9	.93
B × *S*s within groups	4.55	72	—
P × B	5.85	3	1.94
D × P × B	2.71	9	.90
S × P × B	2.56	3	.85
S × P × B × D	3.81	9	1.26
P × B × *S*s within groups	3.02	72	—
Total		255	
Alternative analysis of block effects			
Locus (L)	13.67	3	3.27*
L × D	9.62	9	2.30*
L × S	7.44	3	1.78
L × D × S	7.09	9	1.70
L × *S*s within groups	4.18	72	—
L × P	9.23	3	3.71*
D × P × L	4.00	9	1.61
S × P × L	8.65	3	3.47*
S × P × L × D	2.45	9	.98
L × P × *S*s within groups	2.49	72	—
Total		255	

* $p < .05$.

show that, for the systematic stimuli alone, more errors occurred in the up and down quadrants than in the left and right quadrants when the systematic stimuli were presented before the nonsystematic stimuli. Total mean errors per block for the systematic stimuli when these occurred first in the sequence are: upper quadrant, 12.63; lower quadrant, 12.50; right quadrant, 10.31; left quadrant, 10.94. Differences do not obtain within the systematic stimuli when their presentation is preceded by the nonsystematic presentations (upper, 9.31; lower, 9.81; right, 9.25; left, 10.38).

For the nonsystematic stimuli alone, locus differences did not occur. Quadrant values of the nonsystematic stimuli when they were preceded by the systematic were: upper, 13.88; lower, 13.44; right, 12.88; left, 13.13. When the nonsystematic presentations occurred first, quadrant values were: upper, 13.25; lower, 11.88; right, 12.38; left, 11.75.

D. Discussion and Conclusions

The results of the present investigation, which show an increase in visual acuity with an increase in exposure time, corroborate the findings of previous investigators who have treated the acuity response as a purely retinal event governed by such factors as intensity and duration (1). In addition, the findings demonstrate a differential effect of exposure time on peripheral visual acuity as the result of a systematic and nonsystematic presentation of the stimuli. Significantly more errors occur under conditions of nonsystematic presentation when exposure times are .1 and .2 seconds; with exposure times of .5 and 1 seconds these differences disappear. Apparently, prior knowledge of where the stimulus will appear elicits behavior which enables *S* to improve the accuracy of his perceptual response.

The most parsimonious explanation of the effects of mode of stimulus presentation is in terms of the superiority of central over peripheral vision. It appears likely that, despite instructions to maintain fixation on the central fixation spot, shifts in fixation do indeed occur with the result that "peripheral" stimulus targets are sometimes observed in the central retina. When *S* had prior knowledge of where the stimulus would appear, as was the case during the systematic presentations, a fixation shift could be directed toward the known locus. Such directed shifts could not occur during the nonsystematic presentations and exploratory fixations would be impossible during exposure times as short as .1 and .2 seconds.

The fixation-shift hypothesis receives support from the results of the systematic and nonsystematic presentations when exposure times were .5 and 1 seconds. The shifts could occur as with the short exposures. But with the longer exposure times the nonsystematic stimuli could be located and viewed centrally because sufficient time was available to explore the visual field.

Despite the contention that peripheral stimuli are viewed centrally, the peripheral curves of Figure 1 show more errors at all durations than does the foveal curve. This is so because the probability that fixation shifts bring the stimuli under central regard is never as great during the peripheral measurements as when the stimuli are foveally presented. This probability increases as time increases so that the peripheral curves approach the foveal curve at the longer durations. Duration, per se, of course remains a critical determinant of the acuity response.

The occurrence of quadrant effects among the systematically presented stimuli take the form of fewer errors in the left and right quadrants than in the upper and lower quadrants. Shifts in fixation apparently take place more frequently in the horizontal than in the vertical meridian probably because of well-established reading habits. However, these differences ensue only when the systematic stimuli are preceded by the nonsystematic presentations. Prior experience in detecting nonsystematically presented targets appears to abolish the facilitative tendency to scan horizontally.

The main conclusion to be drawn from these experiments is that, when called upon to make a visual acuity response, *S* will resort to behaviors which experience has shown will yield a maximum resolution of visual detail. Thus, the superior resolving power of the central over the peripheral retina causes *S* to view "peripheral" stimuli centrally, and prior knowledge of stimulus loci facilitates the perceptual act.

REFERENCES

1. GRAHAM, C. H., & COOK, C. Visual acuity as a function of intensity and exposure time. *Amer. J. Psychol.,* 1937, **49**, 654-661.
2. KARN, H. W., & GREGG, L. W. Acquisition of perceptual responses as a function of loading, location, and repetition. *J. Exp. Psychol.,* 1961, **62**, 62-69.

Department of Psychology
Carnegie Institute of Technology
Pittsburgh 13, Pennsylvania